

# Multiplicative Prior and Sequential Monte Carlo Methods for Multiple Gaussian Graphical Models

Linda S. L. Tan\*, Ajay Jasra\*, Maria De Iorio<sup>†</sup> and Timothy M. D. Ebbels<sup>‡</sup>

email: statsll@nus.edu.sg; staja@nus.edu.sg; m.deiorio@ucl.ac.uk; t.ebbels@imperial.ac.uk

\* Department of Statistics and Applied Probability, National University of Singapore

<sup>†</sup> Department of Statistical Science, University College London

<sup>‡</sup> Department of Surgery and Cancer, Imperial College London

## Abstract

We propose using the multiplicative (or Chung-Lu random graph) model as a prior on graphs for Bayesian inference of Gaussian graphical models (GGMs). In the multiplicative model, each edge is chosen independently with probability equal to the product of the connectivities of the end nodes. This class of prior is parsimonious yet highly flexible; it can be used to encourage sparsity or graphs with a pre-specified degree distribution when such prior knowledge is available. We further extend the multiplicative model to multiple GGMs linking the probability of edge inclusion through logistic regression and demonstrate how this leads to joint Bayesian inference for multiple GGMs. A sequential Monte Carlo (SMC) algorithm is developed for estimating the posterior distribution of the graphs. We illustrate the performance of the proposed methods on simulated data and an application which investigates the effect of cadmium (a toxic environmental pollutant) on the correlation structure of a number of urinary metabolites.

**Key Words:** Gaussian Graphical Models; Prior Specification; Sequential Monte Carlo.

## 1 Introduction

Gaussian graphical models (GGMs, Dempster, 1972) provide an important tool for studying the dependence structure among a set of random variables. Under the assumption that the variables have a joint Gaussian distribution, a zero in the precision matrix indicates conditional independence between the associated variables. This corresponds to the absence of an edge in the underlying graph, where nodes denote variables and edges represent conditional dependencies (Lauritzen, 1996). GGMs are widely used, for instance, in biological networks to study the dependence structure among genes from expression data (e.g. Dobra *et al.*, 2004; Chun *et al.*, 2014) and financial time series for forecasting and predictive portfolio analysis

(e.g. Carvalho and West, 2007; Wang *et al.*, 2011). In applications where the effect of different experimental conditions on the dependence relationships among variables is of interest, multiple GGMs (one for each condition) have to be estimated. Under such circumstances, joint inference can encourage sharing of information across graphs and allow for common structure where appropriate (e.g. Guo *et al.*, 2011; Peterson *et al.*, 2015).

We focus on Bayesian inference for GGMs using the G-Wishart prior (Roverato, 2002; Atay-Kayis and Massam, 2005). The G-Wishart is the family of conjugate distributions for the precision matrix, where entries corresponding to missing edges in the underlying graph are constrained to be zero. The normalizing constant of the G-Wishart can only be computed in closed form for decomposable graphs. In this work, we consider the unrestricted graph space where non-decomposable graphs are allowed. Where necessary, we use the Monte Carlo method of Atay-Kayis and Massam (2005) and the Laplace approximation of Lenkoski and Dobra (2011) to estimate the normalizing constant efficiently.

This article makes two main contributions. First, we propose using the multiplicative model of Chung and Lu (2002) as a prior on graphs for estimating GGMs. This prior is further extended to multiple GGMs via logistic regression. Second, we develop a novel sequential Monte Carlo (SMC) algorithm (Del Moral *et al.*, 2006) which uses tempering techniques to obtain joint posterior inference for multiple GGMs. The performance of proposed methods is illustrated on simulated data and an application to urinary metabolic data.

## 1.1 Background and Structure

In the absence of any prior belief on the graphical structure, a uniform prior over all graphs is often used in estimating GGMs (e.g. Lenkoski and Dobra, 2011; Wang and Li, 2012). That is, given  $p$  nodes, it is assumed that each of the  $2^r$  possible graphs, where  $r = p(p-1)/2$ , has equal probability of arising. This prior concentrates its mass on graphs with moderately large number of edges and the expected number of edges as well as the mode is  $r/2$  (see Figure 1). Thus, this prior may not be appropriate when sparse graphs are desired. Several alternatives have been developed. To encourage sparse graphs, Dobra *et al.* (2004) and Jones *et al.* (2005) propose a prior where every edge is included independently with a small probability  $\alpha$  so that a graph with  $x$  edges has prior probability  $\alpha^x(1-\alpha)^{r-x}$ . This prior is known as the Erdős-Rényi model in random graph theory and it reduces to the uniform prior when  $\alpha = 0.5$ . Jones *et al.* (2005) recommend taking  $\alpha = 2/(p-1)$  so that the expected number of edges is  $p$ . Carvalho and Scott (2009) treat  $\alpha$  as a model parameter rather than a fixed tuning constant. They place a Beta( $a, b$ ) prior on  $\alpha$  so that a graph with  $x$  edges

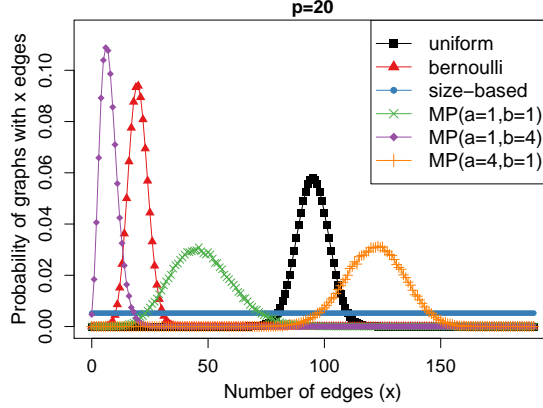


Figure 1: Plot shows the probability allocated to graphs with  $x$  edges by the uniform prior, the Bernoulli prior (Jones *et al.*, 2005) with probability of inclusion of each edge:  $\alpha = 2/(p - 1)$ , the size-based prior (Armstrong *et al.*, 2009) or equivalently the Bernoulli prior with  $\alpha \sim \text{Uniform}[0, 1]$  integrated out (Carvalho and Scott, 2009) and the multiplicative prior (MP) for different values of  $a$  and  $b$ .

has prior probability  $B(a + x, r + b - x)/B(a, b)$ , where  $B(a, b)$  denotes the Beta function. When  $a = b = 1$ , this probability simplifies to  $\frac{1}{r+1} \binom{r}{x}^{-1}$ . This prior is equivalent to the size-based prior (Armstrong *et al.*, 2009) when the graph space is unrestricted. Under the size-based prior, every size  $0, \dots, r$ , has equal probability and every graph of the same size has equal probability. Carvalho and Scott (2009) demonstrate that their proposed prior has strong control over the number of spurious edges and corrects for multiple hypothesis testing automatically, where each null hypothesis corresponds to the exclusion of one edge.

We propose using the multiplicative or Chung-Lu random graph model as a prior on the graphical space of GGMs. Given a desired or expected degree sequence  $\{d_1, \dots, d_p\}$ , where  $d_i$  denotes the degree (number of neighbours) of node  $i$ , the multiplicative model (Chung and Lu, 2002) assumes that the edge between each pair of nodes  $i$  and  $j$  is formed independently with probability  $p_{ij}$  proportional to the product  $d_i d_j$ . Allowing self-loops and provided  $(\max_i d_i)^2 < \sum_i d_i$ , the expected degree of node  $i$  is exactly  $d_i$ . The multiplicative model is able to capture degree distributions which are more diverse (e.g. right-skewed, U-shaped) and closer to that of real-world networks than the Erdős-Rényi model. Notably, the Erdős-Rényi model has a degree distribution that is binomial and can be viewed as a special case of the multiplicative model with a constant expected degree sequence. We consider an alternative parametrization of the multiplicative model introduced by Olhede and Wolfe (2013), which dispenses with self-loops and the normalization constraint by taking  $p_{ij} = \pi_i \pi_j$  and  $0 < \pi_i < 1$  for each  $i$ . They derive degree characteristics and large-sample approximations of this model, which lends insight on the variation attainable in degree structure. Adopting

a Bayesian approach, we treat each  $\pi_i$  as a variable with a  $\text{Beta}(a, b)$  prior. We present degree and clustering properties of the multiplicative model, showing how they depend on choices of  $a$  and  $b$ . In the context of GGMs, we show that the multiplicative model provides an avenue to encourage sparsity or graphs that exhibit particular degree patterns based on prior knowledge obtained through expert opinion or past data. We further demonstrate how the multiplicative model can be extended to include covariates and become a prior on joint graphs for multiple GGMs.

Several approaches for joint inference of multiple GGMs have been developed recently. Guo *et al.* (2011) estimate precision matrices for different groups jointly by parameterizing each entry as a product of two factors: one factor is common across all groups while the other is group specific. A hierarchical  $l_1$  penalty is imposed and optimization is performed using graphical lasso (Friedman *et al.*, 2008). Danaher *et al.* (2014) formulate a more general framework called joint graphical lasso and introduce two convex penalty functions; the fused graphical lasso which encourages edge value on top of structural similarity and the group lasso which only encourages a shared sparsity pattern. Chun *et al.* (2014) extend the approach of Guo *et al.* (2011) to a wider class of nonconvex penalty functions. Yajima *et al.* (2015) compare the Gaussian directed acyclic graphs of two subgroups using Bayesian inference via Gibbs sampling. The strength of association between two variables in the differential group is modeled as the strength in the baseline group plus an edge-specific parameter controlling the difference in association between the two subgroups. They define a prior on the graphical space by centering on a prior graph constructed from a database (Telesca *et al.*, 2012). Peterson *et al.* (2015) consider an alternative Bayesian approach which links graphs from different groups using a Markov Random Field. The probability of inclusion of each edge in graphs  $1, \dots, K$ , is parameterized in terms of a  $K \times K$  symmetric matrix which measures the pairwise similarity of groups and is common across all edges, and an edge-specific  $K \times 1$  vector which controls the inclusion probability in each group independently of group relatedness. Priors are further placed on these parameters and a block Gibbs sampler is used for inference.

The approach that we use to link multiple graphs is based on the multiplicative model by expressing the connectivity of each node as a logistic regression function of graph specific covariates. As the multiplicative model decouples the inclusion probability of each edge into the product of the connectivities of the end nodes, the resulting model is parsimonious and scales linearly in the number of variables and graphs. For inference, we develop an SMC sampler for estimating the joint posterior distribution of the graphs. Using tempering techniques (see e.g. Del Moral *et al.*, 2006, and the references therein), we create a sequence

of probability distributions from which to sample, moving gradually from a distribution that is easy to sample from, through artificial intermediate distributions towards the posterior distribution of interest.

In Section 2, we introduce the multiplicative model and discuss its degree and clustering properties. Section 3 specifies the model setup for multiple GGMs. Section 4 describes posterior inference and a Laplace approximation for the prior probabilities of graphs. The SMC algorithm is outlined in Section 5. Proposed methods are illustrated using simulations and an application to urinary metabolic data in Section 6. Section 7 concludes.

## 2 Multiplicative model

Here we define our notation and present the multiplicative model, followed by a study of its properties. These properties lend insight on the structure and range of networks that can be generated from the multiplicative model. They are also useful in the determination of suitable hyperparameters based on prior understanding of data obtained through expert opinion or a database.

Let  $G = (V, E)$  be a simple graph with vertex set  $V = \{1, 2, \dots, p\}$  and edge set  $E \subseteq \{(i, j) \in V \times V : i < j\}$ . A simple graph is undirected and does not contain self-loops or multiple edges. The adjacency matrix  $A = [A_{ij}]$  of  $G$  is a  $p \times p$  binary matrix where  $A_{ij}$  is 1 if an edge is present between nodes  $i$  and  $j$ , and 0 otherwise for  $i, j \in V$ . As  $G$  is simple,  $A$  is symmetric and has zeros on its diagonal.

In the multiplicative model, each edge is modeled independently as

$$\begin{aligned} A_{ij} &\sim \text{Bernoulli}(p_{ij}) \text{ for } 1 \leq i < j \leq p, \\ p_{ij} &= \pi_i \pi_j, \text{ where } 0 \leq \pi_i \leq 1 \text{ for } i = 1, \dots, p. \end{aligned} \tag{1}$$

Thus, every edge  $A_{ij}$  is formed independently with probability  $p_{ij}$ , where  $p_{ij}$  is a product of the tendencies of nodes  $i$  and  $j$  to form edges with other nodes. The parameter  $\pi_i$  is characteristic of node  $i$  and reflects its activity level. We refer to  $\pi_i$  as the *connectivity* of  $i$ . The Erdős-Rényi random graph model arises as a special case when  $\pi_i$  is constant across all  $i$ , that is, every link is formed independently with equal probability.

We adopt a Bayesian approach and place an independent Beta prior on each  $\pi_i$ . Let

$$\pi_i \sim \text{Beta}(a, b) \text{ for } i = 1, \dots, p, \tag{2}$$

where  $a, b > 0$ . We have  $p(\pi_i) = \pi_i^{a-1}(1 - \pi_i)^{b-1}/B(a, b)$ , where the Beta function  $B(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ . Let  $\pi = (\pi_1, \dots, \pi_p)^T$  and  $p(\pi) = \prod_{i=1}^p p(\pi_i)$ . Networks of highly varying densities and structures can be formed by choosing different hyperparameters  $a$  and  $b$ .

## 2.1 Degree and clustering properties

The degree  $D_i$  of a node  $i$  is the number of links that involve  $i$  or the number of neighbours of  $i$ , and is given by  $D_i = \sum_{j \neq i} A_{ij}$ . The properties below describe the degree distribution and cohesiveness of networks generated from the multiplicative model. Their implications are discussed later in the section. We follow the framework in Rastelli *et al.* (2015), which is based on probability generating functions (see Newman *et al.*, 2001). Proofs are given in the Supplementary Material. We note that some of these results have been discussed in Olhede and Wolfe (2013) but not with regards to the Beta( $a, b$ ) prior. In the following, let  $\mu = \frac{a}{a+b}$  and  $\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)}$  denote the mean and variance of a Beta( $a, b$ ) distribution respectively.

**P1:** The probability that a randomly chosen node is a neighbour of a node with connectivity  $\pi_i$  is  $\mu\pi_i$ .

**P2:** The degree of a node with connectivity  $\pi_i$  is distributed as Binomial( $p-1, \mu\pi_i$ ). Hence its average degree is  $(p-1)\mu\pi_i$ , which is proportional to  $\pi_i$ .

**P3:** The probability generating function of the degree of a randomly chosen node is given by

$$G_{D_i}(z) = \sum_{d=1}^{p-1} P(D_i = d)z^d = \int_0^1 (1 - \mu\pi_i + \mu\pi_i z)^{p-1} p(\pi_i) d\pi_i. \quad (3)$$

The  $k$ th factorial moment of  $D_i$ ,  $E\{D_i(D_i - 1) \dots (D_i - k + 1)\}$ , is given by

$$E\left(\frac{D_i!}{(D_i - k)!}\right) = G_{D_i}^{(k)}(1) = \frac{(p-1)! B(a+k, b)}{(p-1-k)! B(a, b)} \mu^k \text{ for any positive integer } k.$$

**P4:** The average degree of a randomly chosen node is  $E(D_i) = (p-1)\mu^2$  and the variance is  $\text{Var}(D_i) = (p-1)\mu^2\{1 - \mu^2 + (p-2)\sigma^2\}$ .

**P5:** The degree distribution of a randomly chosen node is given by

$$P(D_i = d) = \binom{p-1}{d} \frac{\mu^d}{B(a, b)} \int_0^1 \pi_i^{a+d-1} (1 - \mu\pi_i)^{p-1-d} (1 - \pi_i)^{b-1} d\pi_i$$

for  $d \in \{0, \dots, p-1\}$ . When  $b = 1$ ,  $P(D_i = d) = \binom{p-1}{d} a B(\mu, a+d, p-d) / \mu^a$ , where  $B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$  is the incomplete Beta function.

- P6:** The dispersion index of the degree distribution is given by  $1 - \frac{a\{a^2+(b+1)a-(p-2)b\}}{(a+b)^2(a+b+1)}$ .
- When  $0 < a < \{\sqrt{b^2 + (4p-6)p+1} - b - 1\}/2$ , the distribution has dispersion index greater than 1 and is over-dispersed.
  - When  $a = \{\sqrt{b^2 + (4p-6)p+1} - b - 1\}/2$ , the distribution has dispersion index 1 (equal to that of a Poisson distribution).
  - When  $a > \{\sqrt{b^2 + (4p-6)p+1} - b - 1\}/2$ , the distribution has dispersion index less than 1 (similar to that of a Binomial distribution) and is under-dispersed.
- P7:** The skewness index or Pearson's moment coefficient of skewness of the degree distribution can be computed as  $\{E(D_i^3) - 3E(D_i)\text{Var}(D_i) - E(D_i)^3\} / \text{Var}(D_i)^{1.5}$ , where  $E(D_i^3) = (p-1)\mu^2\{1 + 3(p-2)(\mu^2 + \sigma^2) + (p-2)(p-3)\mu E(\pi_i^3)\}$  and  $E(\pi_i^3) = \frac{(a+2)(a+1)a}{(a+b+2)(a+b+1)(a+b)}$ .
- P8:** The average degree of the neighbours of a node is independent of the connectivity or degree of that node, and is given by  $1 + (p-2)(\mu^2 + \sigma^2)$ .
- P9:** The global clustering coefficient, which measures the probability that nodes  $j$  and  $k$  are linked given that both nodes are linked to  $i$ , is given by  $\frac{a+1}{a+b+1}$ .

In the Erdős-Rényi model,  $D_i$  is distributed as  $\text{Binomial}(p-1, \alpha)$ , where  $\alpha$  is the probability of inclusion of each edge. When  $\alpha = \mu^2$ , the mean degree of a randomly chosen node in the multiplicative model is equal to that in the Erdős-Rényi model from P4. However, the variance of the degree distribution in the multiplicative model is greater than that in the Erdős-Rényi model by  $(p-1)(p-2)\sigma^2$ . Thus, as the number of nodes increases, the multiplicative model can accommodate greater variation in the degree distribution than in the Erdős-Rényi model by  $\mathcal{O}(p^2)$  given the same mean degree.

Figure 2 shows the degree distributions of the multiplicative model for graphs with  $p = 100$  nodes under different hyperparameter settings. When the degree distributions cannot be computed directly using P5, they are estimated via simulation using  $10^5$  graphs. Degree distributions of a variety of shapes (e.g. right-skewed, U-shaped) can be obtained from the multiplicative model by varying  $a$  and  $b$ .

The dispersion index measures how clustered a distribution is compared to standard statistical models. From P6, the degree distribution is over-dispersed when  $a$  is small and under-dispersed when  $a$  is large. In fact, as  $a \rightarrow \infty$  (and/or  $b \rightarrow \infty$ ),  $\sigma^2 \rightarrow 0$  and each  $\pi_i$

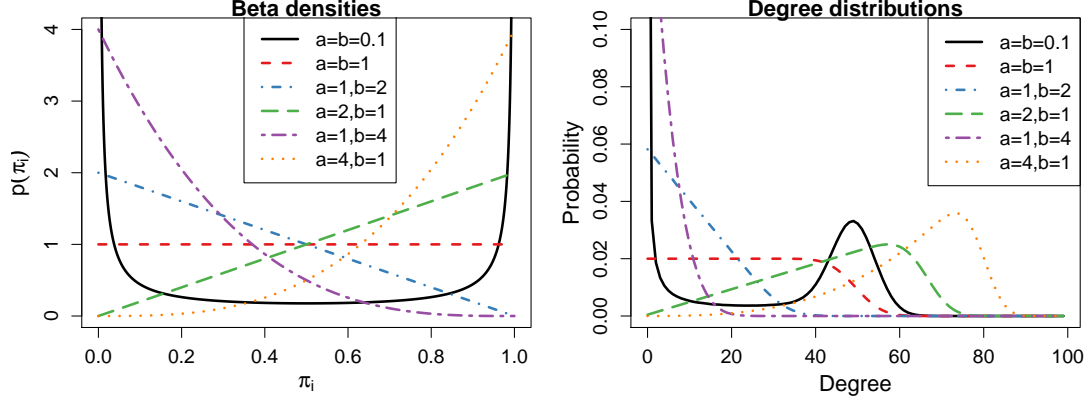


Figure 2: Beta densities (left) and degree distributions of the multiplicative model (right) corresponding to different hyperparameter settings when  $p = 100$ .

reduces to a point mass. The multiplicative model thus degenerates and reduces to the Erdős-Rényi model with constant probability of inclusion for every edge. As the degree distribution is over-dispersed for a wide range of hyperparameter values, the multiplicative model is able to represent well heterogeneity in degree sequences.

The skewness index in P7 is useful for identifying asymmetries in degree distributions. Generally, the degree distribution is positively skewed when  $a$  is small and  $b$  is large and negatively skewed vice versa (plots of the dispersion index and skewness as a function of  $a$  and  $b$  can be found in the Supplementary Material Figures S1 and S2). A network exhibiting power-law behaviour ( $P(D_i = d) \propto d^{-\gamma}$  where  $\gamma$  is a positive constant) tends to have large positive values for the skewness index. For the multiplicative model, Olhede and Wolfe (2013) show that taking  $\pi_i \propto i^{-\gamma}$  for some  $0 < \gamma < 1$  leads to networks with degree distributions that reflect power laws when  $p$  is large. When  $\pi_i$  is given a Beta prior, the skewness is positive and large for small  $a$  and large  $b$  (i.e. sparse networks). In Figure 3, we investigate the relationship between  $\log P(D_i = d)$  and  $\log d$ , which should be a straight line if the power-law is satisfied. We observe that the multiplicative model (with a Beta prior) comes close to power-law behaviour for very sparse networks but the right tail is not sufficiently heavy.

### 3 Gaussian Graphical Models

Suppose we have a dataset with  $p$  variables and  $K$  groups or classes. Let  $y_h = (y_{h1}, \dots, y_{hp})$  denote the  $h^{\text{th}}$  observation of the  $p$  variables for  $h = 1, \dots, H$ , and  $S_k$  be an index set containing the indices of observations which belong to group  $k$  for  $k = 1, \dots, K$ . The number



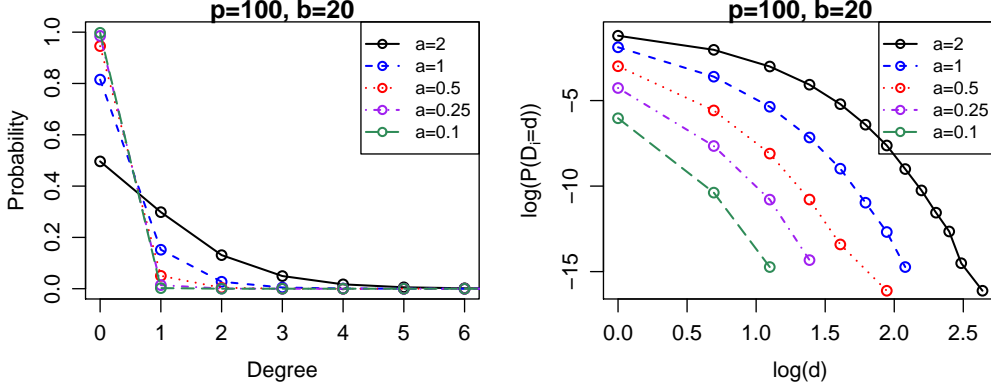


Figure 3: Degree distributions for investigating power-law. Right plot is in log-scale.

of observations in  $S_k$  is denoted by  $|S_k|$  and  $H = \sum_{k=1}^K |S_k|$ . Without loss of generality, we assume that the observations in each group are centered at 0 along each variable. We consider

$$y_h | \Omega_k \sim N(0, \Omega_k^{-1}) \text{ for } h \in S_k, \quad (4)$$

where  $\Omega_k$  is a  $p \times p$  precision matrix and  $k \in \{1, \dots, K\}$ .

Let  $G_k = (V, E_k)$  be a simple graph with vertex set  $V = \{1, 2, \dots, p\}$  and edge set  $E_k \subseteq \{(i, j) \in V \times V : i < j\}$  for  $k = 1, \dots, K$ . Node  $i \in V$  represents the  $i$ th variable and each edge  $(i, j) \in E_k$  corresponds to  $\Omega_{k,ij} \neq 0$ . That is,  $y_{hi}$  and  $y_{hj}$  are conditionally independent (in  $G_k$ ) given the rest of the elements in  $y_h$  if and only if  $\Omega_{k,ij} = 0$ , or equivalently  $(i, j) \notin E_k$ . The conjugate prior for  $\Omega_k$  is the  $G$ -Wishart distribution (Atay-Kayis and Massam, 2005),  $W_{G_k}(\delta_k, D_k)$ , which has density

$$p(\Omega_k | G_k) = \frac{1}{I_{G_k}(\delta_k, D_k)} |\Omega_k|^{(\delta_k - 2)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Omega_k D_k) \right\}.$$

Here,  $\Omega_k$  is constrained to the cone  $P_{G_k}$  of positive definite matrices with entries equal to zero for all  $(i, j) \notin E_k$  and  $I_{G_k}(\delta_k, D_k)$  is a normalizing constant such that

$$I_{G_k}(\delta_k, D_k) = \int_{\Omega_k \in P_{G_k}} |\Omega_k|^{(\delta_k - 2)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Omega_k D_k) \right\} dK.$$

This normalizing constant is guaranteed to be finite if  $\delta_k > 2$  and  $D_k^{-1} \in P_{G_k}$  (Diaconis and Ylvisaker, 1979). The  $G$ -Wishart distribution reduces to the Wishart distribution when  $G_k$

is complete, and the normalizing constant can then be evaluated in closed form as

$$I_{G_k}(\delta_k, D_k) = 2^{(\delta_k + p - 1)p/2} \Gamma_p\{(\delta_k + p - 1)/2\} |D_k|^{-(\delta_k + p - 1)/2}, \quad (5)$$

where  $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=0}^{p-1} \Gamma(a - \frac{1}{2})$  for  $a > (p-1)/2$ .

### 3.1 Priors over graphs

We use the multiplicative model to assign prior probabilities to graphs. Let  $A_k = [A_{k,ij}]$  be the adjacency matrix of  $G_k$  for  $k = 1, \dots, K$ . Consider

$$A_{k,ij} | \pi_{k,i} \pi_{k,j} \sim \text{Bernoulli}(\pi_{k,i} \pi_{k,j}), \quad (6)$$

where  $0 \leq \pi_{k,i} \leq 1$  for  $i = 1, \dots, p$  and  $k = 1, \dots, K$ . As in Section 2, the probability that an edge  $(i, j)$  is present in  $E_k$  is given by  $\pi_{k,i} \pi_{k,j}$ , the product of the propensities of nodes  $i$  and  $j$  to form edges with other nodes in  $G_k$ . Priors are further placed on each  $\pi_{k,i}$ . We consider the cases  $K = 1$  and  $K > 1$  separately.

#### 3.1.1 When $K = 1$

When  $K = 1$ , there is only one group and the subscript  $k$  indicating different groups may be dropped so that  $G_1 = G$  and  $\pi_{1,i} = \pi_i$  for  $i = 1, \dots, p$ . We place a  $\text{Beta}(a, b)$  prior on each  $\pi_i$  as in (2). The prior probability of  $G$  with adjacency matrix  $A$  is then given by

$$\begin{aligned} p(G|a, b) &= \int p(G|\pi) p(\pi|a, b) d\pi \\ &= \frac{1}{B(a, b)^p} \int \prod_{i,j: i < j} (1 - \pi_i \pi_j)^{(1-A_{ij})} \prod_{i=1}^p \pi_i^{(a+d_i-1)} (1 - \pi_i)^{(b-1)} d\pi, \end{aligned} \quad (7)$$

where  $d_i$  denotes the degree of node  $i$ .

#### 3.1.2 When $K > 1$

We propose a joint prior for  $G_1, \dots, G_K$ , which is allowed to depend on covariates specific to each graph. First, we express  $\pi_{k,i}$  in terms of a logistic regression as

$$\pi_{k,i} = \frac{\exp(\beta_i^T x_k)}{1 + \exp(\beta_i^T x_k)}$$

for  $i = 1, \dots, p$ , and  $k = 1, \dots, K$ , where  $x_k = (x_{k1}, \dots, x_{kQ})^T$  is a vector of covariates for  $G_k$  and  $\beta_i = (\beta_{i1}, \dots, \beta_{iQ})$  is a vector of coefficients specific to node  $i$ . Let  $x = (x_1, \dots, x_K)$  and  $\beta = (\beta_1^T, \dots, \beta_p^T)^T$ . We consider a normal prior for each  $\beta_{iq}$  such that

$$\beta_{iq} | \sigma_q^2 \sim N(0, \sigma_q^2)$$

for  $i = 1, \dots, p$  and  $q = 1, \dots, Q$ . Let  $\sigma^2 = (\sigma_1^2, \dots, \sigma_Q^2)$  be a hyperparameter assumed to be known. The joint prior probability of  $G_1, \dots, G_K$ , is then given by

$$\begin{aligned} p(G_1, \dots, G_K | x, \sigma^2) &= \int p(\beta | \sigma^2) \prod_{k=1}^K p(G_k | x_k, \beta) d\beta \\ &= \int \prod_{i=1}^p \prod_{q=1}^Q \left\{ \frac{1}{\sqrt{(2\pi\sigma_q^2)}} \exp\left(-\frac{\beta_{iq}^2}{2\sigma_q^2}\right) \right\} \prod_{k=1}^K \left\{ \prod_{i=1}^p \pi_{k,i}^{d_{k,i}} \prod_{i < j} (1 - \pi_{k,i} \pi_{k,j})^{1-A_{k,ij}} \right\} d\beta, \quad (8) \end{aligned}$$

where  $d_{k,i}$  denotes the degree of node  $i$  in  $G_k$  for  $k = 1, \dots, K$ .

As an example, in the application on urinary metabolic data, we consider  $K = 2$  and the covariates  $x_k$  to include an intercept and an indicator for level of exposure to cadmium (1 if above the median and 0 otherwise). We take  $x_1 = (1, 0)$  and  $x_2 = (1, 1)$  so that  $G_1$  and  $G_2$  represent the correlation structure of the groups with exposure to cadmium below or equal to the median, and above the median respectively. The connectivity of node  $i$  is  $\pi_{1,i} = \{1 + \exp(-\beta_{i1})\}^{-1}$  in  $G_1$  and  $\pi_{2,i} = \{1 + \exp(-\beta_{i1} - \beta_{i2})\}^{-1}$  in  $G_2$ . Thus  $\beta_{i1}$  determines the popularity of node  $i$  in  $G_1$  while  $\beta_{i2}$  is a differential parameter which controls the difference in popularity of node  $i$  between  $G_1$  and  $G_2$ . If  $\beta_{i2}$  is close to zero, the connectivity of node  $i$  in  $G_1$  and  $G_2$  is similar. As the magnitude of  $\beta_{i2}$  increases, the difference in connectivity of node  $i$  between  $G_1$  and  $G_2$  becomes greater. See illustration in Supplementary Material Figure S3.

Figure 4 shows the prior degree distributions of  $G_1$  and  $G_2$  for  $p = 50$  and different values of  $\sigma^2$ . These plots are obtained via simulation of  $10^5$  joint pairs of graphs in each case. When  $\sigma_1^2 = \sigma_2^2 = 0.1$ , both  $\beta_{i1}$  and  $\beta_{i2}$  are close to zero, and  $\pi_{1,i}$  and  $\pi_{2,i}$  are close to 0.5. Thus the degree distribution is shaped like a Binomial curve, resembling the Erdős-Rényi model where each edge is formed independently with constant probability 0.25. As  $\sigma_1^2$  increases, there is greater variation in the degree sequence of  $G_1$ . When  $\sigma_1^2$  is large, the connectivity of each node tends to the extremes of 0 and 1 (each node has a high probability of being either very connected or isolated). Thus the degree distribution resembles the case where each  $\pi_i$  is

allocated a U-shaped Beta(0.1,0.1) prior as shown in Figure 2. The distinction between the degree distribution of  $G_1$  and  $G_2$  becomes greater as  $\sigma_2^2$  increases.

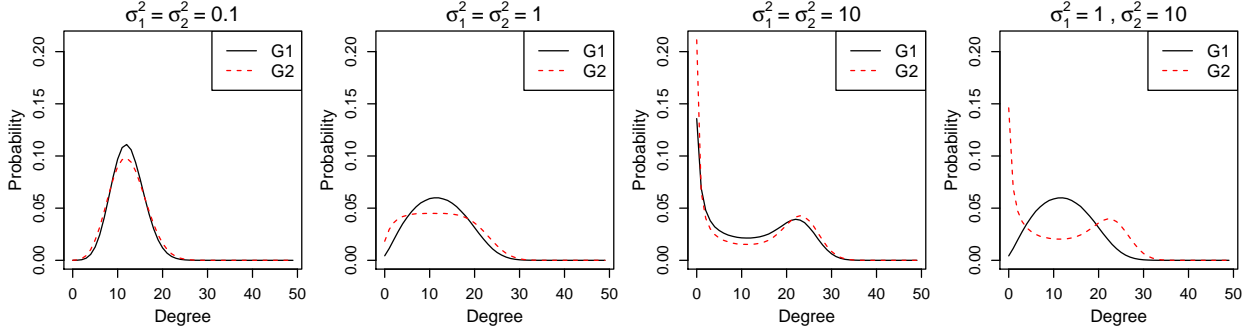


Figure 4: Prior degree distributions of  $G_1$  and  $G_2$  for  $p = 50$  and different combinations of  $\sigma_1^2$  and  $\sigma_2^2$ . Covariates for  $G_1$  and  $G_2$  are (1,0) and (1,1) respectively.

We can also add a third covariate say for gender (1 if male and 0 for female) so that  $K = 4$  and take  $x_1 = (1, 0, 0)$ ,  $x_2 = (1, 0, 1)$ ,  $x_3 = (1, 1, 0)$  and  $x_4 = (1, 1, 1)$ . Then  $G_1$ , for instance, will represent the correlation structure for the group of females with exposure to cadmium below or equal to the median level.

## 4 Posterior distribution

Let  $y = (y_1, \dots, y_H)$ . For  $K > 1$ , the joint distribution of the model can be written as

$$p(y, \Omega_1, \dots, \Omega_K, G_1, \dots, G_K, \beta | x, \sigma^2) = p(\beta | \sigma^2) \prod_{k=1}^K \left\{ p(G_k | x_k, \beta) p(\Omega_k | G_k) \prod_{h \in S_k} p(y_h | \Omega_k) \right\}.$$

Integrating out  $\Omega_k$ , the marginal likelihood  $p(\{y_h | h \in S_k\} | G_k)$  can be shown (see, e.g. Atay-Kayis and Massam, 2005) to be given by

$$p(\{y_h | h \in S_k\} | G_k) = (2\pi)^{-p|S_k|/2} I_{G_k}(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T) / I_{G_k}(\delta_k, D_k).$$

Integrating out  $\beta$  as well, we have

$$p(G_1, \dots, G_K | y, x, \sigma^2) \propto p(G_1, \dots, G_K | x, \sigma^2) \prod_{k=1}^K I_{G_k}(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T) / I_{G_k}(\delta_k, D_k). \quad (9)$$

When  $K = 1$ , the posterior distribution can be derived similarly. The only difference is that the dependence on  $x$  and  $\sigma^2$  is replaced by the Beta prior hyperparameters  $a$  and  $b$ . We have

$$p(G|y, a, b) \propto p(G|a, b) I_G(\delta + H, D_k + \sum_{h=1}^H y_h y_h^T) / I_G(\delta, D). \quad (10)$$

For posterior inference, we propose a SMC algorithm to obtain samples from the posterior distribution. To compute the right-hand side of 9 and 10, we note that for any graph  $G$  (not necessarily decomposable), normalizing constants of the form  $I_G(\delta, D)$  can be evaluated by first factorizing  $G$  into its prime components and their separators (see, e.g. Lauritzen, 1996). Suppose  $(\mathcal{P}_1, \dots, \mathcal{P}_L)$  is a perfect sequence of the prime components of  $G$  and  $(\mathcal{S}_2, \dots, \mathcal{S}_L)$  is the corresponding set of separators. Then  $I_G(\delta, D) = \prod_{l=1}^L I_{G_{\mathcal{P}_l}}(\delta, D) / \prod_{l=2}^L I_{G_{\mathcal{S}_l}}(\delta, D)$ , where  $G_{\mathcal{P}_l}$  and  $G_{\mathcal{S}_l}$  denote the subgraphs induced by  $\mathcal{P}_l$  and  $\mathcal{S}_l$  respectively. As the separators are complete,  $I_{G_{\mathcal{S}_l}}(\delta, D)$  can be evaluated as in (5). The same applies to  $I_{G_{\mathcal{P}_l}}(\delta, D)$  for any prime component  $\mathcal{P}_l$  which is complete. Otherwise, we estimate  $I_{G_{\mathcal{P}_l}}(\delta, D)$  using the Monte Carlo method of Atay-Kayis and Massam (2005) when  $\delta$  is small and the Laplace approximation of Lenkoski and Dobra (2011) when  $\delta$  is large. The priors on graphs are estimated using Laplace approximation, which is described next.

#### 4.1 Laplace approximation for prior on graphs

Evaluating  $p(G|a, b)$  or  $p(G_1, \dots, G_K|x, \sigma^2)$  via Monte Carlo becomes more computationally intensive as  $p$  increases and we estimate these quantities efficiently using Laplace approximation instead. We consider

$$\int \exp\{f(u)\} du \approx (2\pi)^{\frac{n}{2}} | -H(u_0) |^{-\frac{1}{2}} \exp\{f(u_0)\}, \quad (11)$$

where  $u = (u_1, \dots, u_d)^T$ ,  $u_0$  is the mode of  $f(u)$  and  $H(u_0)$  denotes the Hessian of  $f$  evaluated at  $u_0$ . The mode  $u_0$  can be found using numerical methods.

For  $K = 1$ , we estimate  $p(G|a, b)$  in (7) using (11) by first making a change of variable and letting  $\pi_i = \frac{\exp(u_i)}{1 + \exp(u_i)}$ . Then we take

$$\begin{aligned} f(u) = & \sum_i \{(a + d_i)u_i - (a + d_i + 1) \log[1 + \exp(u_i)] + (b - 1) \log(1 - \pi_i)\} \\ & + \sum_{i < j} (1 - A_{ij}) \log(1 - \pi_i \pi_j) - p \log B(a, b). \end{aligned}$$

For  $K > 1$ , we estimate  $p(G_1, \dots, G_K | x, \sigma^2)$  in (8) using (11) by taking  $u = \beta$  and

$$f(\beta) = \sum_{k=1}^K \left\{ \sum_{i=1}^p d_{k,i} \log \pi_{k,i} + \sum_{i < j} (1 - A_{k,ij}) \log(1 - \pi_{k,i} \pi_{k,j}) \right\} - \sum_{i=1}^p \sum_{q=1}^Q \left\{ \frac{1}{2} \log(2\pi\sigma_q^2) + \frac{\beta_{iq}^2}{2\sigma_q^2} \right\}.$$

For simplicity, we have omitted dependence of  $f$  on other variables in its expression. The corresponding gradient and Hessian expressions are given in the Supplementary Material.

## 5 Sequential Monte Carlo sampler

We use SMC samplers for posterior inference. Suppose we are interested in sampling from a complex target  $\lambda(x)$ . The idea is to start with some distribution  $\lambda_1$  that is easy to sample from and move via a sequence of intermediate distributions,  $\lambda_2, \dots, \lambda_{T-1}$ , towards the distribution of interest  $\lambda_T = \lambda$ . At any time  $t$ , a large collection of weighted samples  $\{W_t^{(n)}, X_t^{(n)} | n = 1, \dots, N\}$  is maintained, and these particles are used to generate samples from the target distribution at the next time point using sequential importance sampling (SIS) and resampling methods. The motivation is that it would be easier to move the particles from one target to the next if  $\lambda_t$  is close to  $\lambda_{t+1}$ .

### 5.1 Review of methodology

We first review SIS and SMC briefly. Let  $\lambda_1, \dots, \lambda_T$ , be the target densities and  $\gamma_1, \dots, \gamma_T$ , unnormalized densities such that  $\lambda_t(x_{1:t}) \propto \gamma_t(x_{1:t})$  for  $t = 1, \dots, T$ . Let  $\phi$  be some test function and  $\eta_t$  an arbitrary proposal density. In importance sampling,

$$E_{\lambda_t}[\phi(X_{1:t})] = \int \phi(x_{1:t}) \frac{\lambda_t(x_{1:t})}{\eta_t(x_{1:t})} \eta_t(x_{1:t}) dx_{1:t} = \frac{\int \phi(x_{1:t}) w_t(x_{1:t}) \eta_t(x_{1:t}) dx_{1:t}}{\int w_t(x_{1:t}) \eta_t(x_{1:t}) dx_{1:t}},$$

where

$$w_t(x_{1:t}) = \gamma_t(x_{1:t}) / \eta_t(x_{1:t}) \quad (12)$$

are the unnormalized weights. If  $\{X_{1:t}^{(n)} | n = 1, \dots, N\}$  is a sample from  $\eta_t(x_{1:t})$ , then a Monte Carlo approximation of  $E_{\lambda_t}\{\phi(x_{1:t})\}$  is  $\sum_{n=1}^N \phi(X_{1:t}^{(n)}) W_t^{(n)}$ , where  $w_t^{(n)} = w_t(X_{1:t}^{(n)})$  and

$$W_t^{(n)} = w_t^{(n)} / \sum_{n=1}^N w_t^{(n)} \quad (13)$$

are the normalized weights. Given  $\{W_{1:t}^{(n)}, X_{1:t}^{(n)} | n = 1, \dots, N\}$  approximating  $\lambda_t(x_{1:t})$  at time  $t$ , samples  $\{X_{1:t+1}^{(n)}\}$  approximating  $\lambda_{t+1}$  at time  $t+1$  are obtained in SIS by sampling from  $\{X_{1:t}^{(n)}\}$  using a Markov kernel  $K_{t+1}(x_t, x_{t+1})$ . The proposal density is

$$\eta_{t+1}(x_{1:t+1}) = \eta_t(x_{1:t})K_{t+1}(x_t, x_{t+1}). \quad (14)$$

From (12), the corresponding unnormalized weights can be computed recursively using

$$w_{t+1}(x_{1:t+1}) = \frac{\gamma_{t+1}(x_{1:t+1})}{\eta_{t+1}(x_{1:t+1})} = \frac{\gamma_{t+1}(x_{1:t+1})}{\gamma_t(x_t)K_{t+1}(x_t, x_{t+1})} w_t(x_{1:t}).$$

In SMC, artificial joint target distributions  $\tilde{\lambda}_t(x_{1:t}) \propto \tilde{\gamma}_t(x_{1:t})$  are introduced, where  $\tilde{\gamma}_t(x_{1:t}) = \gamma_t(x_t) \prod_{l=1}^{t-1} L_l(x_{l+1}, x_l)$  and  $L_l(x_{l+1}, x_l)$  is an artificial backward in time Markov kernel. Assume  $\{W_{1:t}^{(n)}, X_{1:t}^{(n)} | n = 1, \dots, N\}$  is a weighted sample approximating  $\tilde{\lambda}_t(x_{1:t})$  at time  $t$  distributed according to  $\eta(x_{1:t})$ . Moving the samples to  $\{X_{1:t+1}^{(n)}\}$  using the Markov kernel in (14), the unnormalized importance weights can be computed as

$$w_{t+1}(x_{1:t+1}) = \tilde{\gamma}_{t+1}(x_{1:t+1})/\eta_{t+1}(x_{1:t+1}) = w_t(x_{1:t})\tilde{w}_{t+1}(x_t, x_{t+1}), \quad (15)$$

where  $\tilde{w}_{t+1}(x_t, x_{t+1}) = \gamma_{t+1}(x_{t+1})L_t(x_{t+1}, x_t)/\{\gamma_t(x_t)K_{t+1}(x_t, x_{t+1})\}$  are unnormalized incremental weights. In the proposed algorithm, we take  $K_{t+1}$  to be an MCMC kernel of invariant distribution  $\lambda_{t+1}$  and  $L_t(x_{t+1}, x_t) = \lambda_{t+1}(x_t)K_{t+1}(x_t, x_{t+1})/\lambda_{t+1}(x_{t+1})$ . See Del Moral *et al.* (2006) Section 3.3.2.3. The unnormalized incremental weights then simplify to

$$\tilde{w}_{t+1}(x_t, x_{t+1}) = \gamma_{t+1}(x_t)/\gamma_t(x_t). \quad (16)$$

## 5.2 Proposed Algorithm

Our aim is to sample from  $p(G_1, \dots, G_K | y, x, \sigma^2)$  in (9) when  $K > 1$  and  $p(G | y, a, b)$  in (10) when  $K = 1$ . Let  $p(G_1, \dots, G_K | y)$  denote the posterior density generally omitting dependence on covariates and hyperparameters. We have  $p(G_1, \dots, G_K | y) \propto \gamma(G_1, \dots, G_K | y)$  where

$$\gamma(G_1, \dots, G_K) = \begin{cases} p(G | a, b) I_G(\delta + H, D + \sum_{h=1}^H y_h y_h^T) / I_G(\delta, D) & \text{if } K = 1, \\ p(G_1, \dots, G_K | x, \sigma^2) \prod_{k=1}^K \frac{I_{G_k}(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T)}{I_{G_k}(\delta_k, D_k)} & \text{if } K > 1. \end{cases}$$

For simplicity, we do not state the dependence of  $\gamma$  on other variables explicitly. To construct the SMC sampler, we devise the following sequence of intermediate target densities,

$$p(G_1, \dots, G_K | y)^{\phi_1}, p(G_1, \dots, G_K | y)^{\phi_2}, \dots, p(G_1, \dots, G_K | y)^{\phi_T},$$

where  $0 < \phi_1 < \phi_2 < \dots < \phi_T = 1$  is a sequence of user-specified temperatures. For greater stability, we use tempering to bridge the target densities so that they evolve smoothly. At each time  $t$ , we maintain  $N$  weighted samples  $\{W_t^{(n)}, (G_1, \dots, G_K)_t^{(n)} | n = 1, \dots, N\}$  approximating the target  $p(G_1, \dots, G_K | y)^{\phi_t} \propto \gamma(G_1, \dots, G_K)^{\phi_t}$  and the annealing temperature is raised gradually from 0 to 1.

### 5.3 Initialization and computation of weights

To generate  $N$  samples from the initial target  $p(G_1, \dots, G_K | y)^{\phi_1}$  at time  $t = 1$ , we sample  $(G_1, \dots, G_K)$  uniformly from the joint graphical space. This can be accomplished by sampling each edge in  $G_k$  independently with probability 0.5 for each  $k = 1, \dots, K$ . This process is performed  $N$  times independently to obtain  $\{(G_1, \dots, G_K)_1^{(n)} | n = 1, \dots, N\}$ . The weight of each sample can be computed using importance sampling. Let  $r = p(p - 1)/2$ . From (12),

$$w_1^{(n)} = \gamma((G_1, \dots, G_K)_1^{(n)})^{\phi_1} 2^{rK}. \quad (17)$$

Suppose we increase the temperature from  $\phi_{t-1}$  to  $\phi_t$  at time  $t \geq 2$ . From (15) and (16), unnormalized weights for the  $n$ th sample can be computed as

$$w_t^{(n)} = w_{(t-1)}^{(n)} \gamma((G_1, \dots, G_K)_{t-1}^{(n)})^{\phi_t - \phi_{t-1}}, \quad (18)$$

Normalized weights may be obtained using (13).

### 5.4 Resampling

To prevent degeneracy of the particle approximation, we measure the effective sample size,  $\text{ESS} = \{\sum_{n=1}^N W_t^{(n)}\}^{-1}$ , at each time  $t$  and resample when the ESS falls below a threshold, say  $N_{\text{threshold}} = N/3$ . Resampling is performed by drawing  $N$  new particles from the multinomial distribution with parameters  $(W_t^{(1)}, \dots, W_t^{(N)})$ . In this way, particles with high weights will be duplicated multiple times while particles with low weights will be eliminated. Resampled particles are then assigned equal weights.



## 5.5 MCMC steps

Suppose we have weighted samples  $\{W_{t-1}^{(n)}, (G_1, \dots, G_K)_{t-1}^{(n)} | n = 1, \dots, N\}$ . At time  $t$ , these samples are moved using an MCMC kernel with invariant distribution  $p_t(G_1, \dots, G_K | y)$  by performing a small number of MCMC steps. This improves mixing and helps to restore the heterogeneity lost during resampling. During this step, candidates for each sample are generated by selecting a small number, say  $M$ , of edges uniformly at random from the set of all possible edges and proposing to flip each edge (a 1 (present) to 0 (absent) and vice versa) in turn in each  $G_k$  for  $k = 1, \dots, K$ . For the selected edge, let  $(G_1, \dots, G_K)_c^{(n)}$  denote the sample obtained after flipping this edge in one of the  $K$  graphs in  $(G_1, \dots, G_K)_{t-1}^{(n)}$ . As the proposal is symmetric, the candidate is accepted with probability given by

$$\min \left[ \left\{ \gamma((G_1, \dots, G_K)_c^{(n)}) / \gamma((G_1, \dots, G_K)_t^{(n)}) \right\}^{\phi_t}, 1 \right]. \quad (19)$$

If the candidate is accepted, we update the  $n$ th sample as  $(G_1, \dots, G_K)_t^{(n)} = (G_1, \dots, G_K)_c^{(n)}$ , otherwise it remains unchanged. The proposed SMC sampler is summarized in Algorithm 1. Note that Algorithm 1 is easily parallelizable as computation of weights as well as the MCMC steps can be performed independently for the  $N$  samples.

---

### Algorithm 1 SMC Algorithm for multiple GGMS

---

At  $t = 1$ ,

- draw  $(G_1, \dots, G_K)_1^{(n)}$  at random uniformly from the joint graphical space for  $n = 1, \dots, N$ .
- Compute weights  $\{w_1^{(n)}\}$  using (17) and obtain normalized weights  $\{W_1^{(n)}\}$  using (13).

For  $t = 2, \dots, T$ ,

- Update weights  $\{w_t^{(n)}\}$  using (18) and obtain normalized weights  $\{W_t^{(n)}\}$  using (13).
  - If  $\text{ESS} < N_{\text{threshold}}$ , resample the particles and set  $W_t^{(n)} = 1/N$  for  $n = 1, \dots, N$ .
  - For  $n = 1, \dots, N$ ,
    - Randomly select  $M$  edges from the set of all possible edges..
    - Set  $(G_1, \dots, G_K)_t^{(n)} = (G_1, \dots, G_K)_{t-1}^{(n)}$ .
    - For  $m = 1, \dots, M$ , and  $k = 1, \dots, K$ , let  $(G_1, \dots, G_K)_c^{(n)}$  be the sample candidate obtained from  $(G_1, \dots, G_K)_t^{(n)}$  by flipping the  $m$ th selected edge in  $G_k$ . Accept sample candidate with probability in (19). If sample candidate is accepted, set  $(G_1, \dots, G_K)_t^{(n)} = (G_1, \dots, G_K)_c^{(n)}$ .
-

## 6 Results

In the following experiments, we take  $\delta_k = 3$  and  $D_k = I_p$  for  $k = 1, \dots, K$  in the G-Wishart prior. For the SMC sampler, we set the number of samples  $N = 500$ , and the number of edges flipped at each iteration in the MCMC step  $M = 5$ . The sequence  $\{\phi_t\}$  is set as  $(0.005, 0.01, \dots, 1)$  with  $T = 200$ . This can also be set adaptively (see Jasra *et al.*, 2011). All code is written in Matlab.

### 6.1 Simulated data

In this section, we investigate the performance of the multiplicative model as a prior on graphs for GGMs and compare it with the commonly used uniform prior which assigns equal prior probability to every graph, and the size-based prior (Armstrong *et al.*, 2009) or equivalently the prior that corrects for multiple hypothesis testing proposed by Carvalho and West (2007) (see Section 1.1 for details). We consider  $p = 20$  nodes and simulate a graph from the multiplicative model in 2 by letting  $\pi_i = 0.9$  for  $i = 1, \dots, 10$ , and  $\pi_i = 0.1$  for  $i = 11, \dots, 20$ , so that each node is either highly or poorly connected. The degree sequence of the simulated graph is  $\{0, 0, 0, 1, 2, 2, 2, 2, 2, 2, 6, 7, 7, 7, 8, 8, 9, 10, 10, 11\}$ . A plot of the simulated graph and its degree distribution can be found in the Supplementary Material Figure S4. Using this graph as the “true” graph, we simulate ten data sets from the GGM in (4) by setting  $H = 200$  and  $K = 1$ . The precision matrix  $\Omega$  is constructed by setting diagonal entries as one and entries corresponding to missing edges as zero. Upper diagonal entries corresponding to edges in  $G$  are simulated randomly from the uniform distribution on  $[-0.4, -0.35] \cup [0.35, 0.4]$ . The resulting matrix is averaged with its transpose to obtain a symmetric matrix that is verified to be positive definite.

Using Algorithm 1, weighted samples from the posterior distribution are obtained for each simulated data set under the uniform prior, size-based prior and multiplicative prior respectively. For the multiplicative model, we set the Beta hyperparameters as  $a = b = 0.1$  (see Figure 2) so that the shape of the prior degree distribution resembles that of the true graph. Using the weighted samples, we compute the posterior probability of the occurrence of each edge, and summarize the results using ROC curves in Figure 5. The mean and standard deviation of the area under the ROC curves (AUC) over the ten data sets are given in Table 1. While there is some variation in results over different simulated data sets, the multiplicative prior generally performs as well as and in several cases better than the uniform and size-based prior, with a higher AUC value on the average. We observe that the multiplicative model

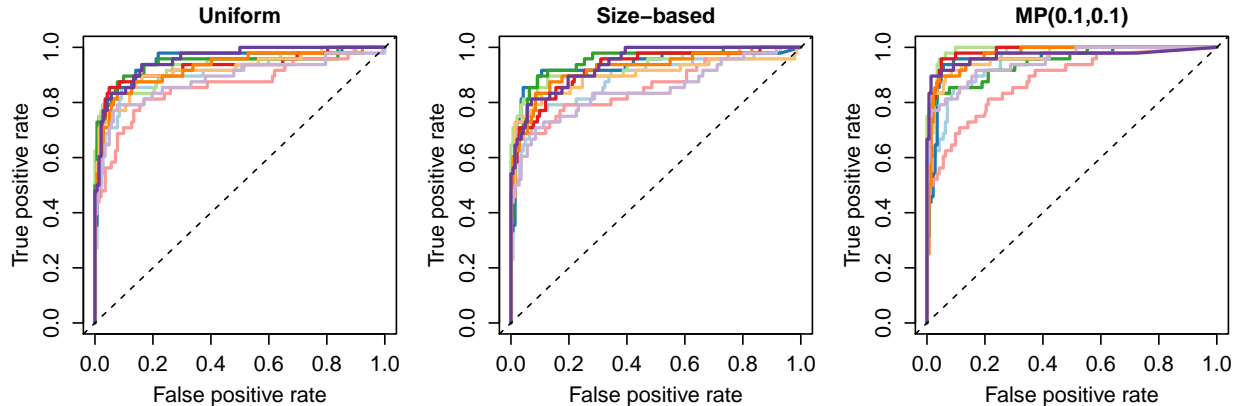


Figure 5: ROC curves corresponding to different priors. Each color corresponds to one of the ten simulated data sets.

	Uniform	Size-based	MP(0.1,0.1)
AUC	0.92 (0.03)	0.91 (0.04)	0.96 (0.03)

Table 1: AUC values corresponding to different priors averaged over ten simulated data sets. Standard deviation given in brackets.

demonstrates the ability to encourage sparsity generally and in particular graphs where the connectivity of each node is close to either 0 or 1.

## 6.2 Application to urinary metabolic data

We analyze urinary metabolic data for  $H = 127$  individuals acquired using  $^1\text{H}$  NMR spectroscopy (see Ellis *et al.* (2012) for details). These individuals live close to a lead and zinc smelter at Avonmouth in Bristol, UK, which produces large quantities of airborne cadmium (Cd). An extremely toxic metal, cadmium is commonly released through industrial processes and acute exposure poses numerous health risks. Here, we investigate the correlation structure of  $p = 22$  urinary metabolites listed in Table 2 in response to cadmium exposure through GGMs. This dataset has also been studied by Salamanca *et al.* (2014) using Bayesian hierarchical models. We perform two analyses. In the first case, we consider the individuals as a homogeneous group. In the second case, we divide the individuals into two groups;  $S_1$  (a control group with level of exposure to cadmium lower than or equal to the median) and  $S_2$  (a diseased group with level of cadmium higher than the median). In each case, we first use the R package **GeneNet** (Schaefer *et al.*, 2015) to obtain fast shrinkage estimators of partial correlation in the network. The degree distributions obtained (see Supplementary Material Figure S5) can be used as a basis for determining appropriate hyperparameters for the mul-

multiplicative model. The observations of each variable are first normalized to have zero mean and standard deviation of one in each group.

Metabolites	Abbrev	Degree				Betweenness			
		M(1, 1)	M(0.1, 0.1)	SB	UF	M(1, 1)	M(0.1, 0.1)	SB	UF
Trimethylamine oxide	TMAO	<b>4.52</b>	<b>4.62</b>	<b>4.20</b>	5.33	<b>0.15</b>	<b>0.13</b>	<b>0.20</b>	0.08
P-cresol-sulphate	PCS	4.33	3.81	3.72	<b>6.39</b>	0.10	0.07	0.11	0.09
Succinate	Suc	3.95	4.07	3.79	5.71	0.10	0.10	0.17	0.07
Dimethylamine	DMA	3.81	4.05	3.47	5.68	0.10	0.08	0.12	0.07
Creatinine	Creat	3.54	3.38	3.09	4.93	0.08	0.05	0.09	0.06
4-deoxyerythronic acid	4-DEA	3.52	1.56	2.63	5.23	0.09	0.02	0.11	0.08
Pyruvate	Pyr	3.21	3.19	3.01	5.41	0.05	0.05	0.06	0.06
Citrate	Cit	3.15	2.58	2.42	4.84	0.09	0.07	0.12	0.05
3-hydroxyisovalerate	3-HV	2.85	2.87	2.74	4.58	0.10	0.08	0.17	0.06
Glycine	Gly	2.70	2.79	2.41	4.42	0.07	0.06	0.10	0.05
Urea	Urea	2.62	3.50	2.09	6.17	0.07	0.07	0.06	<b>0.09</b>
Alanine	Ala	2.48	2.01	2.38	5.45	0.06	0.03	0.10	0.07
Phenylacetylglutamine	PAG	2.24	2.12	2.16	4.55	0.02	0.02	0.05	0.04
Acetate	AcO	1.93	0.29	1.32	5.13	0.04	0.00	0.04	0.05
Hippurate	Hip	1.55	2.98	1.76	4.86	0.02	0.05	0.04	0.06
Dimethylglycine	DMG	1.49	1.88	1.44	4.70	0.02	0.03	0.04	0.06
Trimethylamine	TMA	1.28	1.18	1.60	3.28	0.01	0.01	0.04	0.03
Lactate	Lac	1.08	0.62	0.96	4.56	0.02	0.01	0.02	0.04
Proline-betaine	PB	0.64	0.06	0.84	2.49	0.00	0.00	0.01	0.01
N-methyl-nicotinic acid	NMNA	0.56	0.09	1.06	3.06	0.01	0.00	0.01	0.02
Formate	For	0.36	0.03	0.43	2.85	0.00	0.00	0.01	0.02
Creatine	Crea	0.12	0.01	0.28	1.94	0.00	0.00	0.00	0.01

Table 2: List of 22 urinary metabolites and their abbreviations. Columns 3–6 and columns 7–10 show the weighted mean degree and betweenness respectively, under the multiplicative model with  $a = b = 1$  (M(1, 1)) and  $a = b = 0.1$  (M(0.1, 0.1)), the size-based prior (SB) and the uniform prior (UF). The highest value in each column (3–10) is highlighted in bold.

### 6.3 Case: $K = 1$

In this section, we study the correlation structure of the metabolites treating the individuals as one homogeneous group. We compare the performance of four priors on the graphical space, namely, the multiplicative model with  $a = b = 1$  and  $a = b = 0.1$ , the size-based prior and the uniform prior. We fit a GGM to the data using Algorithm 1, obtaining  $N = 500$  weighted samples from the posterior distribution in each case. Using these weighted samples, we compute the posterior probability of occurrence of each edge. Figure 6 shows the graphs obtained under each prior. Only edges with posterior probability greater than 0.5 and associated nodes are shown and the width of each edge is proportional to its posterior

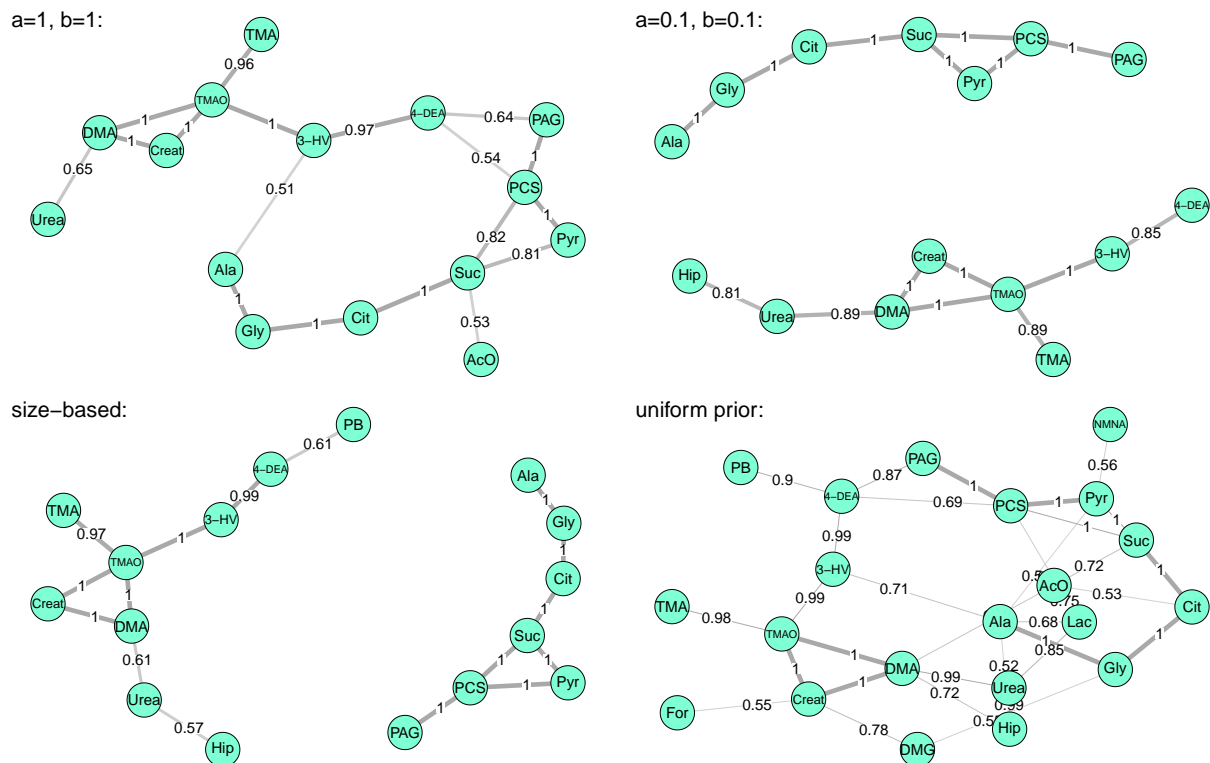


Figure 6: Graphs corresponding to different priors. Only edges with posterior weights greater than 0.5 are shown.

probability. Graphs showing the full set of nodes and all possible edges are given in the Supplementary Material Figure S6. The graphs obtained under the multiplicative model and the size-based prior have a high degree of similarity and are much sparser than that of the uniform prior.

Table 2 shows the weighted mean degree and betweenness centrality measures for each metabolite. The metabolites have been sorted in terms of weighted mean degree in decreasing order according to  $M(1,1)$ , the multiplicative model with  $a = b = 1$ . Under the multiplicative model and size-based prior, TMAO has the highest degree as well as betweenness. Under the uniform prior, PCS has the highest degree with Urea close behind; these two metabolites also have the highest betweenness.

For the multiplicative model, we can also obtain uncertainty measures of the tendency of each node to form connections with other nodes. Figure 7 shows the posterior distributions of the connectivities  $\pi_i$  of each metabolite obtained via simulations. It appears that the multiplicative model with  $a = b = 0.1$  is too strong and places too much prior weight on values of  $\pi_i$  at the extremes of 0 and 1. The multiplicative model with  $a = b = 1$  provides a

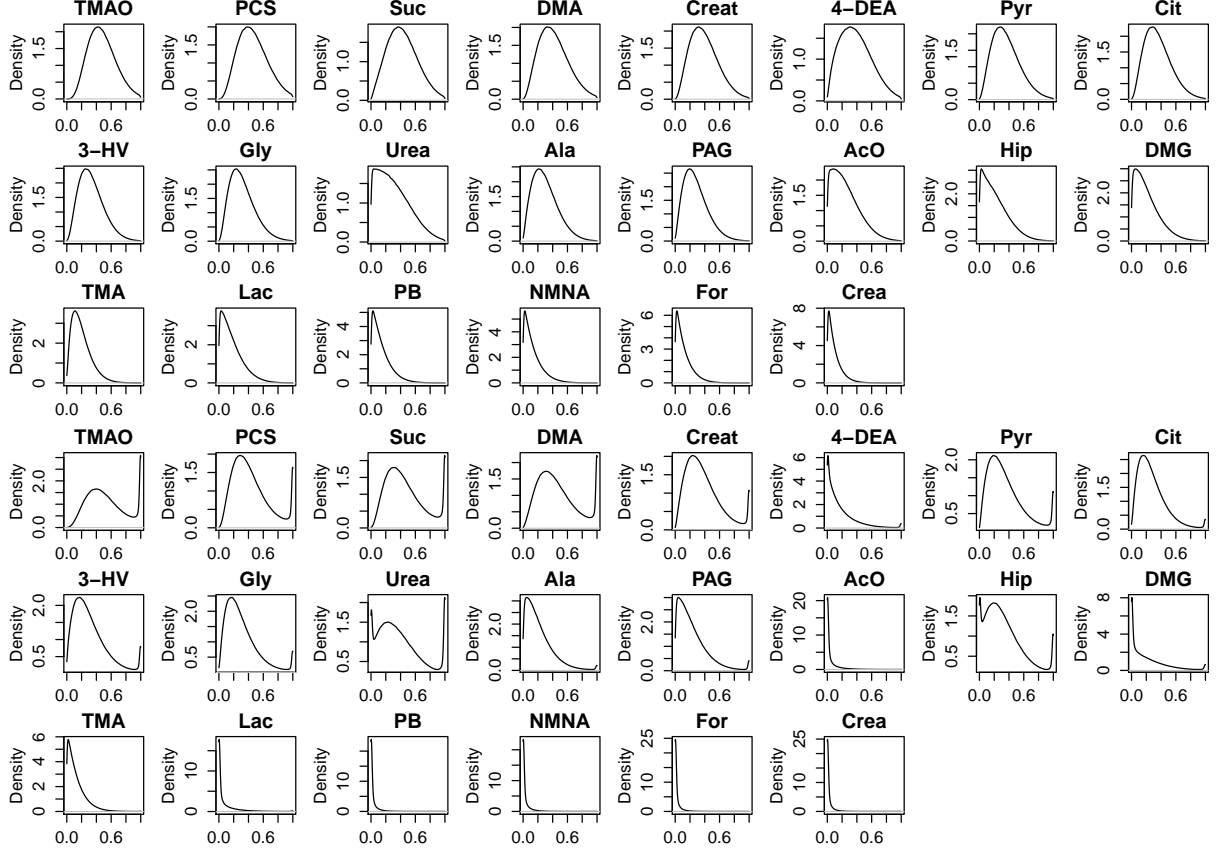


Figure 7: Posterior distribution of the connectivity ( $\pi_i$ ) of different metabolites under the multiplicative model with  $a = b = 1$  (first 3 rows) and  $a = b = 0.1$  (last 3 rows).

better fit. The mean and 95% credible interval of the connectivity of each metabolite, and the mean covariance matrix corresponding to the multiplicative model with  $a = b = 1$  are given in the Supplementary Material Tables S1 and S2.

#### 6.4 Case: $K = 2$

Next, we investigate the difference in correlation structure of the urinary metabolites between the two groups of individuals  $S_1$  (with level of exposure to cadmium lower than or equal to the median) and  $S_2$  (level of exposure higher than the median). We consider the covariates  $x_k$  for the  $k$ th graph to include an intercept and an indicator for level of exposure to cadmium (1 if above the median and 0 otherwise) so that  $x_1 = (1, 0)$  and  $x_2 = (1, 1)$ . The difference in graphical structure between  $G_1$  and  $G_2$  due to exposure to urinary cadmium is of interest. We fit a GGM with  $K = 2$  to the data using the SMC algorithm under four priors. The first three are the multiplicative model with  $\sigma_1^2 = \sigma_2^2 = 1$ ,  $\sigma_1^2 = 1$  and  $\sigma_2^2 = 10$  and  $\sigma_1^2 = \sigma_2^2 = 10$ ,

and the last is the uniform prior. From Figure 4 and the preliminary degree distributions obtained using GeneNet (see Supplementary Material Figure S5), taking  $\sigma_1^2 = \sigma_2^2 = 1$  seems appropriate but we wish to investigate if there is any benefit to be gained by allowing the structure of  $G_2$  to vary more significantly from that of  $G_1$  by taking  $\sigma_2^2$  to be 10 and whether a prior which assumes the tendencies to connect are closer to the extremes of 0 and 1 is more appropriate ( $\sigma_1^2 = \sigma_2^2 = 10$ ).

Using Algorithm 1, we obtain weighted samples from the posterior distribution under each of the four priors. The ESS and acceptance rate in the SMC sampler are monitored at each iteration and these plots are given in the Supplementary Material Figure S7 for the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ . Typically, the ESS decreases as the algorithm proceeds until it falls below the threshold,  $N_{\text{threshold}} = N/3$ , and it bounces back after resampling is performed. Due to bridging of target densities using tempering, the acceptance rate is usually high at the beginning when the temperature  $\phi_t$  is close to zero and proposals have a high probability of being accepted [see (19)]. As the temperature increases, the samples becomes more concentrated in the regions of high posterior probability and the acceptance rate falls.

Figure 8 shows the differential network corresponding to the different priors. It highlights the differences between  $G_1$  and  $G_2$  by displaying only edges which have high posterior probability (e.g. greater than 0.5) of appearing in one graph but not the other (see, e.g. Valcarcel *et al.*, 2011). Due to space limitations, further detailed results are given in the Supplementary Material. These include weighted graphs obtained from Algorithm 1 under different priors (Figures S8 and S9), posterior distributions (Figures S10, S11 and S12), betweenness centrality measures, weighted means and 95% credible intervals of the connectivities ( $\pi_{i,k}$ ) and regression coefficients ( $\beta_{iq}$ ) of each metabolite (Tables S4, S5 and S6) and the mean covariance matrices corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$  (Tables S7 and S8).

The full network in the  $K = 1$  case and the differential network in the  $K = 2$  case both show similar topological characteristics corresponding to sub-graphs of metabolites. For the case of  $K = 2$ , the different prior hyperparameters only lead to different levels of shrinkage, but there is a high degree of similarity in terms of biological interpretation. For example, both Figures 6 and 8 show three different sub-graphs linking metabolites with shared metabolic origin. First, a group of organic acids including succinate, pyruvate, acetate and para-cresol sulphate (PCS) are connected, sometimes also with phenylacetylglutamine (PAG). Several of these metabolites (PCS and PAG) are known to be of gut bacterial origin, and Cd stress is known to modulate gut microbiota populations in mice (Liu *et al.*, 2014).

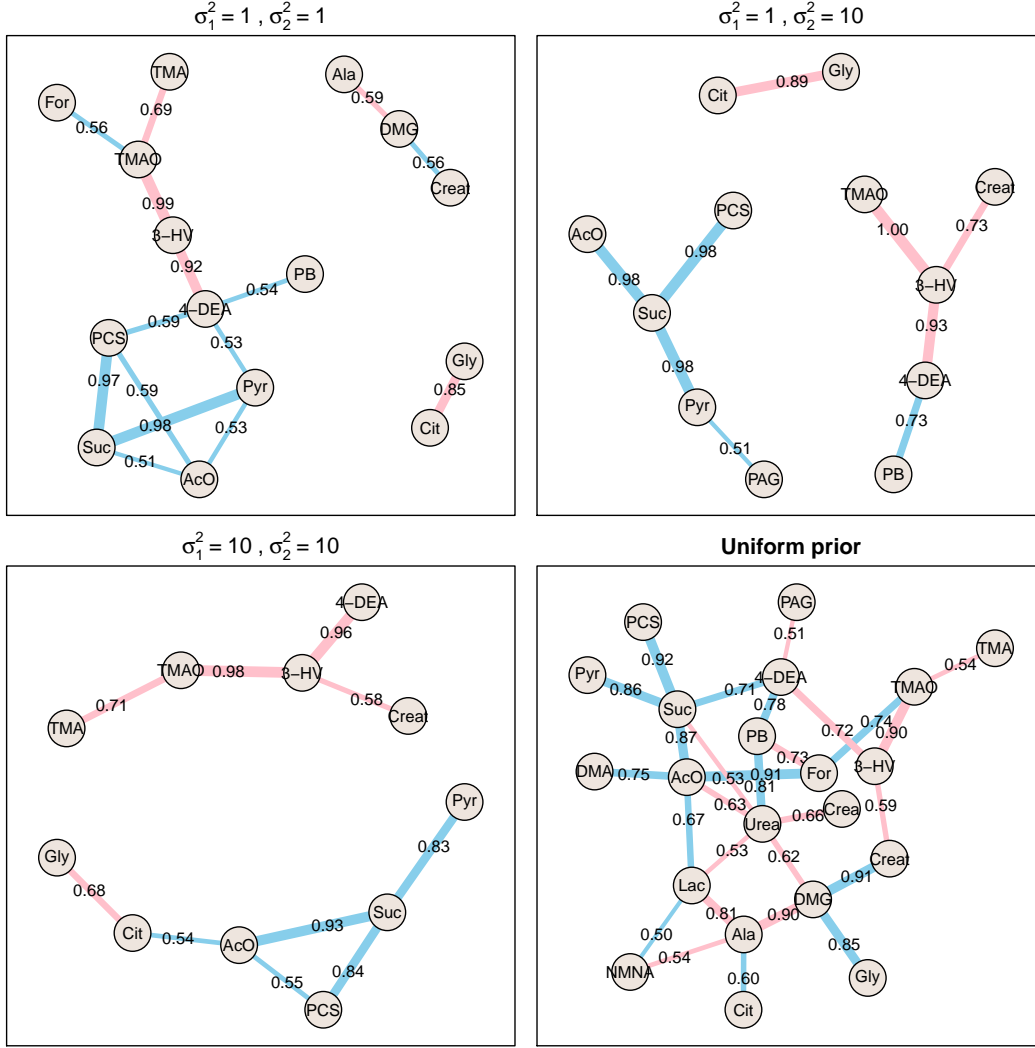


Figure 8: Differential network corresponding to the different priors. Blue edges represent edges with posterior probability more than 0.5 of appearing in  $G_1$  but not in  $G_2$  and pink edges represent edges with posterior probability more than 0.5 of appearing in  $G_2$  but not in  $G_1$ .

Increased acetate is a known consequence of renal damage, which could be linked to high Cd levels in this population. The second group contains trimethylamine (TMA) and its oxidation product trimethylamine-N-oxide (TMAO), both part of choline metabolism, plus 3-HV and 4-DEA which are products of amino acid catabolism. Choline is an essential nutrient and is metabolised primarily in the liver. Due to its long biological half-life, Cd accumulates in human tissues, especially the liver and kidney, so this observation may point towards a possible mechanistic connection. Moreover, in their original study of this data set, Ellis *et al.* (2012) reported a positive correlation between urinary Cd and both 4-DEA and 3-HV, though



this relationship did not survive correction for age and sex. The third group links citrate and glycine, closely associated via malate and glyoxylate in central carbon metabolism (the network of metabolic reactions essential to life). A strong correlation between Cd and citrate was found by Ellis *et al.* (2012), while Valcarcel *et al.* (2011) found a significant deregulation of the dependency network associated with dimethylglycine, a biproduct of the synthesis of glycine from choline. Thus, it is plausible that several of the metabolites found in the networks of Figures 6 and 8 are involved in pathways disregulated due to Cd exposure. However, metabolite associations derive from a variety of factors and many may be indirect, and possibly non-biochemical in origin, e.g. change in expression of membrane transporters. Thus interpretation of dependency networks, such as those generated here, is difficult. Nonetheless, they give us a novel view of the data not exposed in conventional analyses, and may serve to help generate new hypotheses to be investigated by future biochemical experiments.

## 7 Conclusion

This article proposes using the multiplicative or Chung-Lu random graph model as a prior on the graphical space of GGMs, where the probability of inclusion of each edge is a product of the connectivities of the end nodes. This model can be used to encourage sparsity or particular degree structures, when such prior knowledge is available, say from a database or based on expert opinion. A Bayesian approach is adopted and priors are further placed on the connectivity of the nodes. We study the degree and clustering properties of the multiplicative prior and note that this prior is able to accommodate a wider range of degree structures than the Erdős-Rényi model. For example, we can use it to encourage shrinkage towards the extremes of 0 and 1, or degree distributions that are right-skewed by varying the hyperparameters. We illustrate how this prior can be applied to both single and multiple GGMs and a SMC sampler is developed for posterior inference. We find the performance of this sampler to be stable and consistent in our experiments and it can also be parallelized easily. The multiplicative prior also yields rich posterior inference, enabling a study of the connectivity of each node and how the propensity to connect varies across different experimental conditions in the case of multiple GGMs. This allows deeper exploration into the structure of dependency networks and may aid in the formulation of new scientific hypothesis and in opening further lines of investigations.

## 8 Acknowledgments

Linda Tan is supported by the National University of Singapore Overseas Postdoctoral Fellowship. Ajay Jasra is supported by AcRF Tier 1 grant R-155-000-156-112.

## References

- Armstrong, H., Carter, C. K., Wong, K. F. K. and Kohn, R. (2009). Bayesian covariance matrix estimation using a mixture of decomposable graphical models. *Statistics and Computing*, 19, 303–316.
- Atay-Kayis, A. and Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, 92, 317–335.
- Carvalho C. M. and Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika*, 3, 497–512.
- Carvalho, C. M. and West, M. (2007). Dynamic Matrix-Variate Graphical Models. *Bayesian Analysis*, 2, 69–98.
- Chun, H., Zhang, X. and Zhao, H. (2014). Gene regulation network inference with joint sparse Gaussian graphical models. *Journal of Computational and Graphical Statistics*, DOI: 10.1080/10618600.2014.956876.
- Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 15879–15882.
- Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B*, 76, 373–397.
- Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Statistical Royal Society: Series B*, 68, 411–436.
- Dempster, A. P. (1972). Covariance selection. *Biometrics*, 157–175.
- Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics*, 7, 269–281.
- Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G. and West, M. (2004). Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90, 196–212.
- Ellis, J. K., Athersuch, T. J., Thomas, L. D., Teichert, F., Perez-Trujillo, M., Svendsen, C., Spurgeon, D. J., Singh, R., Jarup, L., Bundy, J. G. and Keun, H. C. (2012). Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Medicine*, 10:61. doi:10.1186/1741-7015-10-61.

- Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9, 432–441.
- Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98, 1–15.
- Jasra, A., Stephens, D. A., Doucet, A. and Tsagaris, T. (2011). Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo *Scandinavian Journal of Statistics*, 38, 1–22.
- Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C. and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20, 388–400.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
- Lenkoski, A. and Dobra, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *Journal of Computational and Graphical Statistics*, 20, 140–157.
- Liu Y., Li, Y., Liu, K. and Shen J. (2014). Exposing to cadmium stress cause profound toxic effect on microbiota of the mice intestinal tract. *PLoS ONE*, 9(2): e85323. doi:10.1371/journal.pone.0085323
- Newman, M. E. J., Strogatz, S. H. and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64. doi: 10.1103/PhysRevE.64.026118.
- Olhede, S. C. and Wolfe, P. J. (2013). Degree-based network models. arXiv:1211.6537
- Peterson, C., Stingo, F. C. and Vannucci, M. (2015). Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*, 110, 159–174.
- Rastelli, R., Friel, N. and Raftery, A. E. (2015). Properties of latent variable network models. arXiv:1506.07806.
- Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, 29, 391–411.
- Salamanca, B. V., Ebbels, T. M. D. and De Iorio, M. (2014). Variance and covariance heterogeneity analysis for detection of metabolites associated with cadmium exposure. *Statistical Applications in Genetic and Molecular Biology*
- Schaefer, J., Opgen-Rhein, R. and Strimmer, K. (2015). R package: GeneNet version 1.2.13. <https://cran.r-project.org/web/packages/GeneNet/index.html>
- Telescar, D., Müller, P., Parmigiani, G. and Freedman, R. S. (2012). Modeling dependent gene expression. *The Annals of Applied Statistics*, 6, 542560.
- Valcárcel, B., Würtz, P., Seich al Basatena N.-K., Tukiainen, T., Kangas, A. J., Soininen, P., Järvelin, M.-R., Ala-Korpela, M., Ebbels, T. M. and de Iorio, M. (2011) A differential network approach to exploring differences between biological states: an application to prediabetes. *PLoS ONE*, 6(9): e24702. doi:10.1371/journal.pone.0024702

- Wang, H., Reesony, C. and Carvalho, C. M. (2011). Dynamic financial index models: modeling conditional dependencies via graphs. *Bayesian Analysis*, 6, 639–664.
- Wang, H. and Li, S. Z. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics*, 6, 168–198.
- Yajima, M., Telesca, D. and Müller, P. (2015). Detecting differential patterns of interaction in molecular pathways. *Biostatistics*, 16, 240–251

# Supplementary Material for “Multiplicative Prior and Sequential Monte Carlo Methods for Multiple Gaussian Graphical Models”

Linda S. L. Tan\*, Ajay Jasra\*, Maria De Iorio<sup>†</sup> and Timothy M. D. Ebbels<sup>‡</sup>

\* Department of Statistics and Applied Probability, National University of Singapore

<sup>†</sup> Department of Statistical Science, University College London

<sup>‡</sup> Department of Surgery and Cancer, Imperial College London

## S1 Properties of multiplicative model

### S1.1 Derivations

This section provides the proofs of properties P1 - P9 given in Section 2.1 of the manuscript.

- Proof of P1: The probability of a random node  $j$  being connected to a node  $i$  with connectivity  $\pi_i$  is given by  $\int_0^1 \pi_i \pi_j p(\pi_j) d\pi_j = \pi_i \mu$ .
- Proof of P2: From P1,  $A_{ij}|\pi_i \sim \text{Bernoulli}(\pi_i \mu)$  for any  $j \neq i$ . Since  $D_i = \sum_{j \neq i} A_{ij}$ ,  $D_i|\pi_i \sim \text{Binomial}(p-1, \pi_i \mu)$ . Therefore  $E(D_i|\pi_i) = (p-1)\mu\pi_i$ . We also have  $E(D_i^2|\pi_i) = (p-1)\mu\pi_i + (p-1)(p-2)\mu^2\pi_i^2$ .
- Proof of P3:

$$\begin{aligned} E(z^{D_i}|\pi_i) &= \sum_{d=0}^{p-1} Pr(D_i = d|\pi_i) z^d \\ &= \sum_{d=0}^{p-1} \binom{p-1}{d} (\mu\pi_i z)^d (1 - \mu\pi_i)^{p-1-d} \\ &= (1 - \mu\pi_i + \mu\pi_i z)^{p-1}. \end{aligned}$$

Therefore  $E(z^{D_i}) = E\{E(z^{D_i}|\pi_i)\} = \int_0^1 (1 - \mu\pi_i + \mu\pi_i z)^{p-1} p(\pi_i) d\pi_i$ .

$$\begin{aligned} G_{D_i}^{(k)}(z) &= \int_0^1 p(\pi_i) \frac{\partial^k}{\partial z^k} (1 - \mu\pi_i + \mu\pi_i z)^{p-1} d\pi_i \\ &= \int_0^1 p(\pi_i) \frac{(p-1)!}{(p-1-k)!} (\mu\pi_i)^k (1 - \mu\pi_i + \mu\pi_i z)^{p-1} d\pi_i. \end{aligned}$$

Hence  $G_{D_i}^{(k)}(1) = \frac{(p-1)!}{(p-1-k)!} \mu^k \int_0^1 p(\pi_i) \pi_i^k d\pi_i = \frac{(p-1)! B(a+k, b)}{(p-1-k)! B(a, b)} \mu^k$ .

- Proof of P4: We have

$$E(D_i) = E\{E(D_i|\pi_i)\} = (p-1)\mu^2.$$

$$E(D_i^2) = E\{E(D_i^2|\pi_i)\} = (p-1)\mu^2 + (p-1)(p-2)\mu^2(\mu^2 + \sigma^2).$$

$$\text{Hence, } \text{Var}(D_i) = E(D_i^2) - E(D_i)^2 = (p-1)\mu^2\{1 - \mu^2 + (p-2)\sigma^2\}.$$

- Proof of P5: Since  $P(D_i = d|\pi_i) = \binom{p-1}{d}(\mu\pi_i)^d(1 - \mu\pi_i)^{p-1-d}$ , we have  $P(D_i = d) = \int_0^1 P(D_i = d|\pi_i)p(\pi_i)d\pi_i$ .
- Proof of P6: The dispersion index of the degree distribution can be computed as  $\frac{\text{Var}(D_i)}{E(D_i)} = 1 - \mu^2 + (p-2)\sigma^2$ . Substituting  $\mu = \frac{a}{a+b}$  and  $\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)}$ , we get the result.
- Proof of P7:

$$E(D_i^3|\pi_i) = (p-1)(p-2)(p-3)(\mu\pi_i)^3 + 3(p-1)(p-2)(\mu\pi_i)^2 + (p-1)\mu\pi_i.$$

$$\text{Hence } E(D_i^3) = (p-1)(p-2)(p-3)\mu^3 E(\pi_i^3) + 3(p-1)(p-2)\mu^2(\mu^2 + \sigma^2) + (p-1)\mu^2,$$

$$\text{where } E(\pi_i^3) = \frac{(a+2)(a+1)a}{(a+b+2)(a+b+1)(a+b)}.$$

- Proof of P8: The average degree of node  $j$  given that it is connected to a node with connectivity  $\pi_i$  is given by

$$\begin{aligned} & \sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, \pi_i) \\ &= \int \sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, \pi_i, \pi_j) P(\pi_j | A_{ij} = 1, \pi_i) d\pi_j \\ &= \int \sum_{d=1}^{p-1} d \binom{p-2}{d-1} (\pi_j \mu)^{d-1} (1 - \pi_j \mu)^{p-2-(d-1)} \frac{P(A_{ij} = 1 | \pi_i, \pi_j) p(\pi_i) p(\pi_j)}{P(A_{ij} = 1 | \pi_j) p(\pi_i)} d\pi_j \\ &= \int \sum_{x=0}^{p-2} (x+1) \binom{p-2}{x} (\pi_j \mu)^x (1 - \pi_j \mu)^{p-2-x} \frac{\pi_i \pi_j}{\pi_i \mu} p(\pi_j) d\pi_j \\ &= \int \{(p-2)\pi_j \mu + 1\} \frac{\pi_j}{\mu} p(\pi_j) d\pi_j \\ &= (p-2)(\mu^2 + \sigma^2) + 1 \end{aligned}$$

Therefore the average degree of a neighbour of node  $i$  is independent of its connectivity

$\pi_i$ . Similarly, the average degree of node  $j$  given that it is connected to a node with degree  $k$  is given by

$$\begin{aligned}
& \sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, D_i = k) \\
&= \int \sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, D_i = k, \pi_i) P(\pi_i | A_{ij} = 1, D_i = k) d\pi_i \quad (20) \\
&= \int \sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, \pi_i) \frac{P(\pi_i, A_{ij} = 1, D_i = k)}{P(A_{ij} = 1, D_i = k)} d\pi_i.
\end{aligned}$$

From above, we have  $\sum_{d=1}^{p-1} d P(D_j = d | A_{ij} = 1, \pi_i) = (p-2)(\mu^2 + \sigma^2) + 1$  and

$$\begin{aligned}
\frac{P(\pi_i, A_{ij} = 1, D_i = k)}{P(A_{ij} = 1, D_i = k)} &= \frac{P(D_i = k | \pi_i, A_{ij} = 1) P(A_{ij} = 1 | \pi_i) p(\pi_i)}{\int P(D_i = k | \pi_i, A_{ij} = 1) P(A_{ij} = 1 | \pi_i) p(\pi_i) d\pi_i} \\
&= \frac{\binom{p-2}{k-1} (\pi\mu)^k (1 - \pi\mu)^{p-1-k} p(\pi_i)}{\int \binom{p-2}{k-1} (\pi\mu)^k (1 - \pi\mu)^{p-1-k} p(\pi_i) d\pi_i} \\
&= \frac{P(D_i = k | \pi_i) p(\pi_i)}{P(D_i = k)}.
\end{aligned}$$

Therefore (20) simplifies to  $(p-2)(\mu^2 + \sigma^2) + 1$ , which is again independent of the degree of  $i$ .

- Proof of P9: The global clustering coefficient can be written as

$$\frac{\int (\pi_i \pi_k) (\pi_k \pi_j) (\pi_j \pi_i) p(\pi_i) p(\pi_j) p(\pi_k) d\pi_i d\pi_j d\pi_k}{\int (\pi_i \pi_k) (\pi_k \pi_j) p(\pi_i) p(\pi_j) p(\pi_k) d\pi_i d\pi_j d\pi_k} = \frac{(\sigma^2 + \mu^2)^3}{(\sigma^2 + \mu^2)^2 \mu} = \frac{a+1}{a+b+1}.$$

## S1.2 Dispersion index and skewness

Figures S1 and S2 show plots of the dispersion index and skewness as a function of  $a$  and  $b$  for  $p = 100$ . The left plots in Figures S1 and S2 show how the dispersion index and skewness vary across a wide range of hyperparameter values. The right plot in Figure S1 shows some cross-sectional plots of the skewness while the right plot in Figure S2 compares the skewness of the multiplicative prior with the Erdős-Rényi model for the same mean degree when  $p = 100$  and  $b = 5$ . The multiplicative model tends to be more positively skewed for small values of  $a$  and less so for large values of  $a$  as compared to the Erdős-Rényi model.

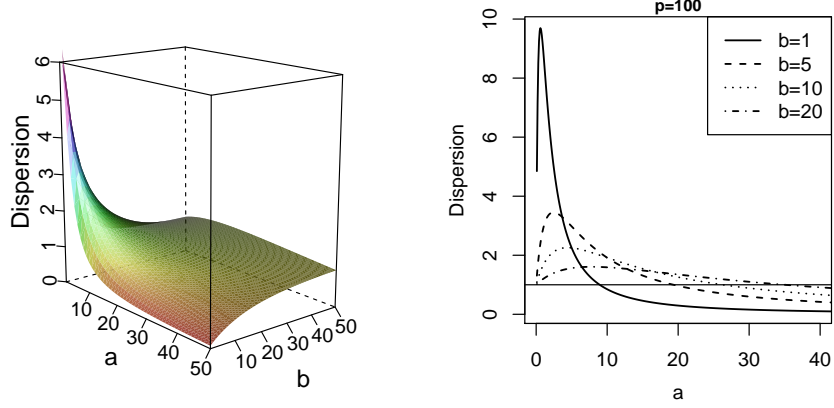


Figure S1: Left: Dispersion index as a function of hyperparameters  $a$  and  $b$  in the Beta prior for  $p = 100$ . Right: Cross-sectional plots for  $b = 1, 5, 10, 20$ .

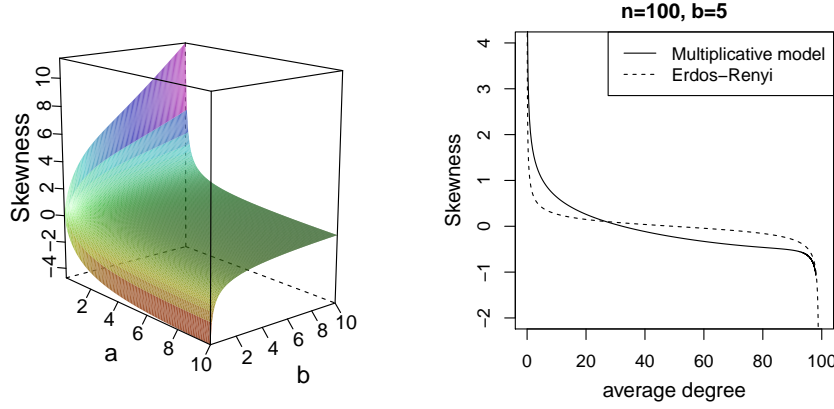


Figure S2: Left: Skewness as a function of hyperparameters  $a$  and  $b$  in the Beta prior for  $p = 100$ . Right: Comparison of skewness of multiplicative prior with the Erdős-Rényi model for the same mean degree.

## S2 Illustration of connectivity of nodes when $K = 2$

In our proposed prior for multiple GGMs, the connectivity of node  $i$  is  $\pi_{1,i} = \{1 + \exp(-\beta_{i1})\}^{-1}$  in  $G_1$  and  $\pi_{2,i} = \{1 + \exp(-\beta_{i1} - \beta_{i2})\}^{-1}$  in  $G_2$  when  $K = 2$ ,  $x_1 = (1, 0)$  and  $x_2 = (1, 1)$ . The relationship between the connectivity of a node and its regression coefficients is illustrated in Figure S3.

## S3 Laplace approximation

The gradients and Hessians required in the Laplace approximation for computing prior probabilities of graphs described in Section 4.1 are given below. They can be derived using vector



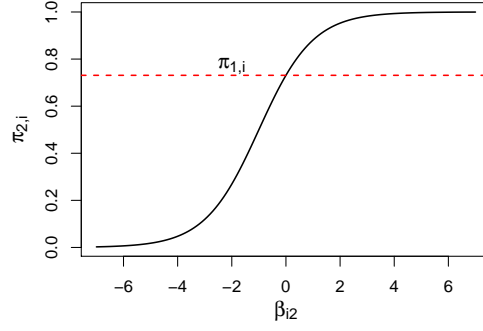


Figure S3: Connectivity of node  $i$  in  $G_2$  ( $\pi_{2,i}$ ) as a function of  $\beta_{i2}$  (black line), with  $\beta_{i1}$  fixed at 1. Red dotted line marks the value of  $\pi_{1,i}$ .

differential calculus and a useful reference is Magnus and Neudecker (1988).

For  $K = 1$ , we have  $\frac{d\pi_i}{du_i} = \frac{\pi_i}{1+\exp(u_i)}$ ,  $\frac{d^2\pi_i}{du_i^2} = \frac{d\pi_i}{du_i} \frac{1-\exp(u_i)}{1+\exp(u_i)}$  and

- $\frac{\partial f}{\partial u_i} = a + d_i - (a + d_i + 1)\pi_i - (b - 1)\frac{d\pi_i}{du_i}/(1 - \pi_i) - \sum_{j \neq i} (1 - A_{ij})\pi_j \frac{d\pi_i}{du_i}/(1 - \pi_i\pi_j),$
- $\frac{\partial^2 f}{\partial u_i \partial u_j} = -(1 - A_{ij})\frac{d\pi_j}{du_j} \frac{d\pi_i}{du_i}/(1 - \pi_i\pi_j)^2,$
- $\frac{\partial^2 f}{\partial u_i^2} = -(a + d_i + 1)\frac{d\pi_i}{du_i} - (b - 1)\left\{ \frac{d^2\pi_i}{du_i^2}/(1 - \pi_i) + \left(\frac{d\pi_i}{du_i}\right)^2/(1 - \pi_i)^2 \right\} - \sum_{j \neq i} (1 - A_{ij})\left\{ \pi_j \frac{d^2\pi_i}{du_i^2}/(1 - \pi_i\pi_j) + \pi_j^2 \left(\frac{d\pi_i}{du_i}\right)^2/(1 - \pi_i\pi_j)^2 \right\}.$

For  $K > 1$ , we have  $\frac{\partial \pi_{k,i}}{\partial \beta_i} = \frac{\pi_{k,i} x_k}{1+\exp(\beta_i^T x_k)}$ ,  $\frac{\partial^2 \pi_{k,i}}{\partial \beta_i \partial \beta_i^T} = \pi_{k,i} \frac{1-\exp(\beta_i^T x_k)}{\{1+\exp(\beta_i^T x_k)\}^2} x_k x_k^T$  and

- $\frac{\partial f}{\partial \beta_i} = \sum_{k=1}^K \left\{ \frac{d_{k,i}}{1+\exp(\beta_i^T x_k)} - \sum_{j \neq i} \frac{(1-A_{k,ij})\pi_{k,i}\pi_{k,j}}{(1-\pi_{k,i}\pi_{k,j})\{1+\exp(\beta_i^T x_k)\}} \right\} x_k - \text{diag}(\frac{1}{\sigma^2})\beta_i,$
- $\frac{\partial^2 f}{\partial \beta_i \partial \beta_j^T} = -\sum_{k=1}^K (1 - A_{k,ij})(1 - \pi_{k,i}\pi_{k,j})^{-2} \frac{\partial \pi_{k,i}}{\partial \beta_i} \left( \frac{\partial \pi_{k,j}}{\partial \beta_j} \right)^T,$
- $\frac{\partial^2 f}{\partial \beta_i \partial \beta_i^T} = -\sum_{k=1}^K \left\{ \frac{d_{k,i}\pi_{k,i}}{1+\exp(\beta_i^T x_k)} + \sum_{j \neq i} \frac{(1-A_{k,ij})\pi_{k,i}\pi_{k,j}\{1-\exp(\beta_i^T x_k)(1-\pi_{k,i}\pi_{k,j})\}}{(1-\pi_{k,i}\pi_{k,j})^2\{1+\exp(\beta_i^T x_k)\}^2} \right\} x_k x_k^T - \text{diag}(\frac{1}{\sigma^2}),$

where  $\frac{1}{\sigma^2} = (\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_Q^2})$  is evaluated element-wise.

## S4 Simulated data

Figure S4 shows the true graph simulated from the multiplicative model in (2) and its degree distribution.

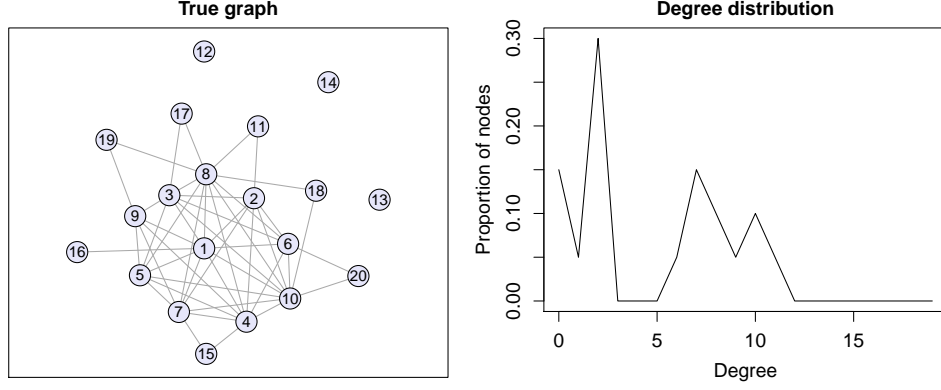


Figure S4: Left: True graph from which data are simulated. Right: Degree distribution of true graph.

## S5 Urinary metabolic data

Figure S5 shows the degree distributions estimated using GeneNet.

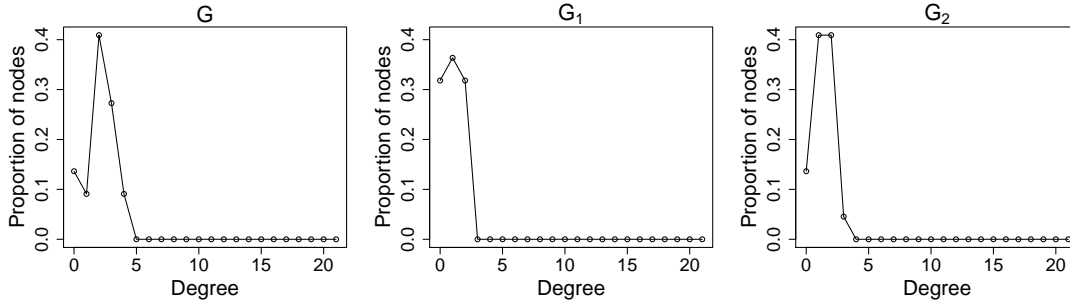


Figure S5: Degree distributions estimated using GeneNet for the case where the individuals are treated as one heterogeneous group (left) and where they are divided into two groups  $S_1$  (middle) and  $S_2$  (right).

### S5.1 Case: $K = 1$

Figure S6 shows the graphs obtained from Algorithm 1 under each prior. The width of every edge is proportional to its posterior weight. Table S1 shows the mean and 95% credible interval of the connectivity of each metabolite. Table S2 shows the mean precision matrix corresponding to the multiplicative prior with  $a = b = 1$ .

### S5.2 Case: $K = 2$

Figure S7 shows a plot of the ESS and mean acceptance rate in the MCMC steps at each iteration for the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ . Figure S8 shows the graphs of  $G_1$  and  $G_2$  corresponding to different priors. The width of the edges are proportional to their

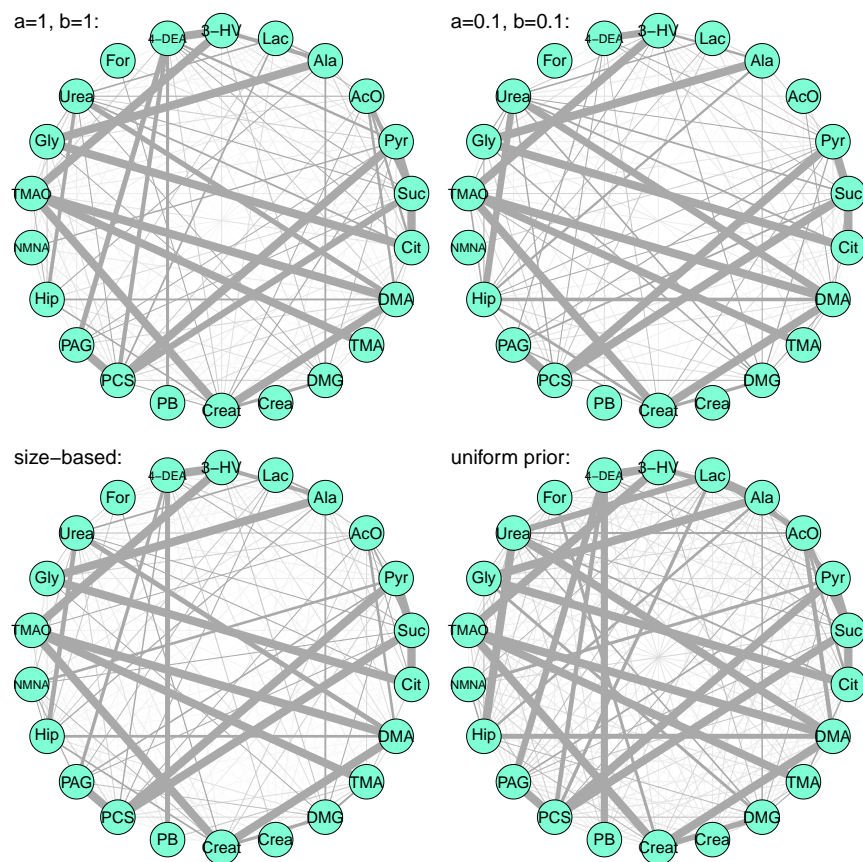


Figure S6: Graphs corresponding to different priors. Width of edges are proportional to their posterior weights.

Abbrev	M(1, 1)		M(0.1, 0.1)	
	$\bar{\pi}_i$	CI ( $\pi_i$ )	$\bar{\pi}_i$	CI ( $\pi_i$ )
TMAO	0.48	(0.15, 0.84)	0.54	(0.00, 0.90)
PCS	0.46	(0.11, 0.84)	0.43	(0.06, 0.96)
Suc	0.43	(0.06, 0.81)	0.46	(0.07, 0.95)
DMA	0.42	(0.07, 0.80)	0.46	(0.07, 0.96)
Creat	0.39	(0.07, 0.76)	0.38	(0.06, 0.99)
4-DEA	0.39	(0.02, 0.79)	0.18	(0.00, 0.14)
Pyr	0.36	(0.05, 0.73)	0.36	(0.05, 0.99)
Cit	0.36	(0.05, 0.71)	0.28	(0.00, 0.65)
3-HV	0.33	(0.05, 0.66)	0.32	(0.00, 0.81)
Gly	0.32	(0.04, 0.66)	0.31	(0.00, 0.76)
Urea	0.31	(0.00, 0.75)	0.40	(0.01, 1.00)
Ala	0.30	(0.02, 0.63)	0.22	(0.00, 0.62)
PAG	0.28	(0.02, 0.60)	0.23	(0.00, 0.68)
AcO	0.25	(0.00, 0.63)	0.05	(0.00, 0.05)
Hip	0.22	(0.00, 0.60)	0.33	(0.04, 1.00)
DMG	0.22	(0.00, 0.58)	0.21	(0.00, 0.19)
TMA	0.20	(0.00, 0.46)	0.13	(0.00, 0.11)
Lac	0.18	(0.00, 0.52)	0.08	(0.00, 0.07)
PB	0.14	(0.00, 0.47)	0.02	(0.00, 0.03)
NMNA	0.14	(0.00, 0.48)	0.03	(0.00, 0.03)
For	0.12	(0.00, 0.44)	0.02	(0.00, 0.03)
Crea	0.10	(0.00, 0.43)	0.02	(0.00, 0.03)

Table S1: Mean and 95% credible interval of the connectivity ( $\pi_i$ ) of each metabolite under the multiplicative prior with  $a = b = 1$  (M(1, 1)) and  $a = b = 0.1$  (M(0.1, 0.1)).

	TMAO	PCS	Suc	DMA	Creat	4-DEA	Pyr	Cit	3-HV	Gly	Urea	Ala	PAG	AcO	Hip	DMG	TMA	Lac	PB	NMNA	For	Crea
TMAO	2.64	0.00	0.00	-1.86	0.82	0.00	0.00	0.00	-0.52	-0.00	-0.01	-0.00	-0.00	0.00	0.02	-0.01	-0.43	-0.00	-0.00	-0.00	-0.00	-0.00
PCS	.	5.93	-0.82	0.00	0.02	-0.36	-3.62	-0.00	0.00	0.01	-0.02	-0.00	-2.11	-0.02	0.00	0.01	-0.00	0.06	0.00	0.01	0.00	0.00
Suc	.	.	1.53	0.00	-0.00	-0.04	0.67	-0.57	-0.00	-0.00	0.05	-0.00	0.05	-0.19	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.01	-0.00
DMA	.	.	.	3.02	-1.30	-0.01	-0.00	-0.00	-0.00	-0.00	0.24	-0.00	-0.00	0.10	0.09	-0.01	-0.02	0.00	-0.00	0.00	0.00	-0.00
Creat	.	.	.	.	1.71	-0.01	0.02	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.02	0.06	-0.12	-0.00	0.00	0.00	0.01	0.03	0.00
4-DEA	.	.	.	.	.	1.34	0.15	0.00	-0.41	-0.00	0.00	-0.00	0.36	-0.00	0.00	-0.00	-0.00	-0.00	-0.13	0.00	0.00	-0.00
Pyr	.	.	.	.	.	4.23	-0.00	0.00	-0.00	-0.01	-0.02	-0.08	0.01	0.00	0.02	-0.00	0.01	0.00	0.06	0.00	0.00	0.00
Cit	.	.	.	.	.	.	1.54	-0.00	-0.52	0.07	-0.01	-0.00	0.11	0.00	0.00	0.02	0.00	0.00	0.00	-0.00	0.00	0.00
3-HV	.	.	.	.	.	.	.	1.46	-0.00	0.00	-0.15	0.00	0.00	0.00	-0.02	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00
Gly	.	.	.	.	.	.	.	.	1.56	0.03	-0.60	0.00	0.01	0.02	-0.05	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00
Urea	.	.	.	.	.	.	.	.	.	1.21	0.00	-0.00	-0.04	0.17	-0.03	-0.00	0.06	0.00	0.00	0.00	-0.00	-0.00
Ala	.	.	.	.	.	.	.	.	.	1.41	0.00	0.00	0.05	-0.07	-0.00	-0.06	0.00	0.00	-0.00	-0.00	-0.00	-0.00
PAG	.	.	.	.	.	.	.	.	.	.	2.83	0.00	-0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	-0.00	-0.00
AcO	.	.	.	.	.	.	.	.	.	.	.	1.14	0.00	0.00	0.00	0.00	-0.02	0.00	0.00	-0.00	-0.00	-0.00
Hip	.	.	.	.	.	.	.	.	.	.	.	.	1.13	0.01	0.00	0.00	-0.00	-0.01	-0.00	0.00	0.00	0.00
DMG	.	.	.	.	.	.	.	.	.	.	.	.	.	1.11	-0.01	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
TMA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.21	0.00	0.00	-0.00	-0.00	0.00	0.00	0.00
Lac	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.08	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
PB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.07	0.00	-0.00	0.00	0.00	0.00
NMNA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.05	-0.00	0.00	0.00	0.00
For	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.04	0.00	0.00	0.00
Crea	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.03	0.00	0.00

Table S2: Mean precision matrix  $\Omega$  corresponding to  $K = 1$  and multiplicative prior with  $a = b = 1$ .

posterior weights. Figure S9 shows these graphs but displaying only edges with posterior weights greater than 0.5 and associated nodes. Table S3 shows a list of 10 edges which are most likely to appear in  $G_1$  but not in  $G_2$  and vice versa under each prior. Tables S4, S5 and S6 show the mean and 95% credible interval of (1) the connectivity ( $\pi_{ik}$ ) and (2) the regression coefficients ( $\beta_{iq}$ ) of each metabolite for  $k = 1, 2$ ,  $q = 1, 2$ , and the weighted mean betweenness centrality measure in each graph under the multiplicative priors. Figures S10, S11 and S12 show the posterior distributions of the connectivity and regression coefficients of each metabolite under the multiplicative priors. Tables S7 and S8 show the mean precision matrix of  $G_1$  and  $G_2$  corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .

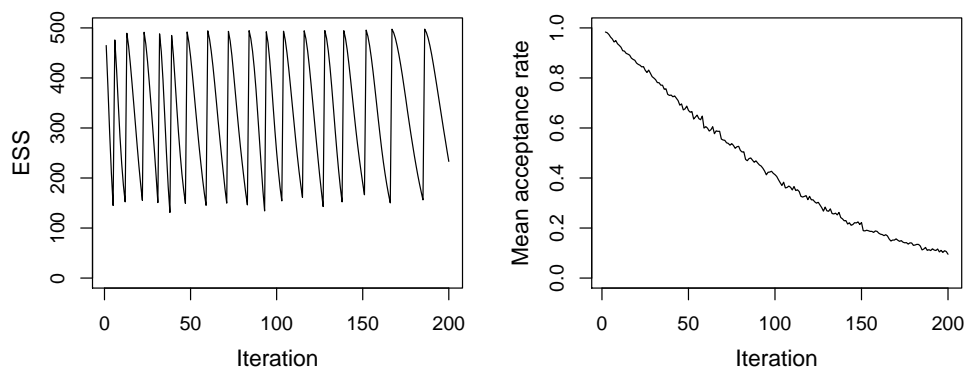


Figure S7: Typical plot of ESS (left) and mean acceptance rate (right) of SMC algorithm. This plot is obtained from fitting the urinary metabolic data using  $K = 2$  for the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .

## References

Magnus, J. R. and Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics. Wiley, Chichester, UK.

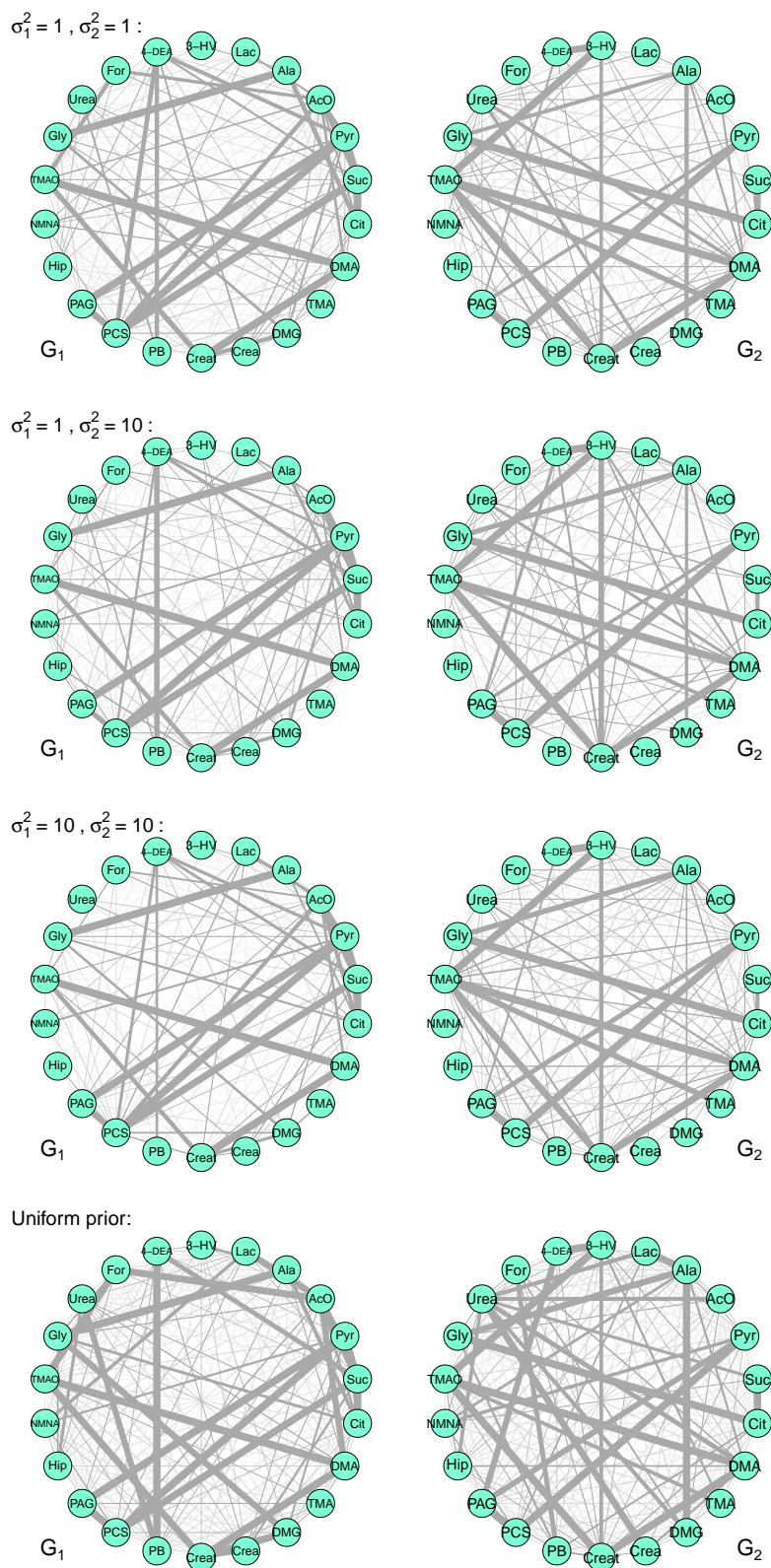
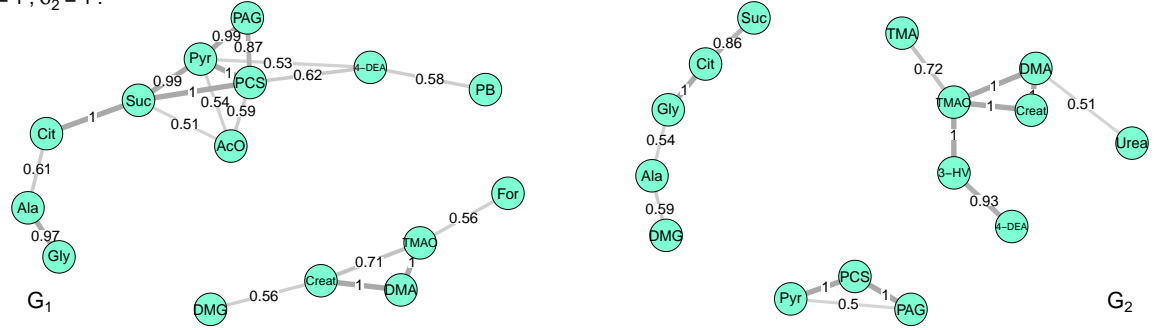
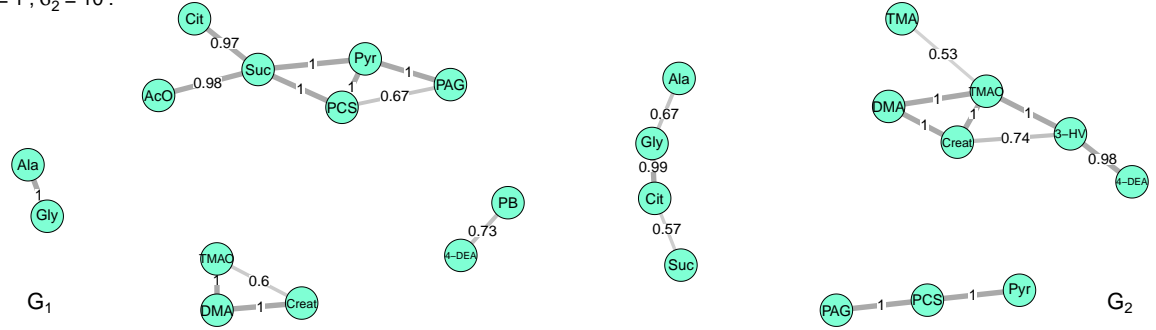


Figure S8: Graphs of  $G_1$  and  $G_2$  corresponding to different priors. Width of edges are proportional to their posterior weights.

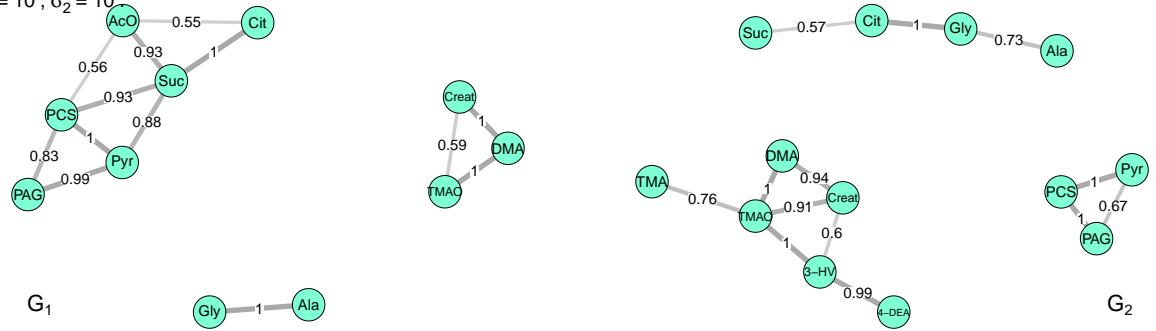
$$\sigma_1^2 = 1, \sigma_2^2 = 1 :$$



$$\sigma_1^2 = 1, \sigma_2^2 = 10 :$$



$$\sigma_1^2 = 10, \sigma_2^2 = 10 :$$



Uniform prior:

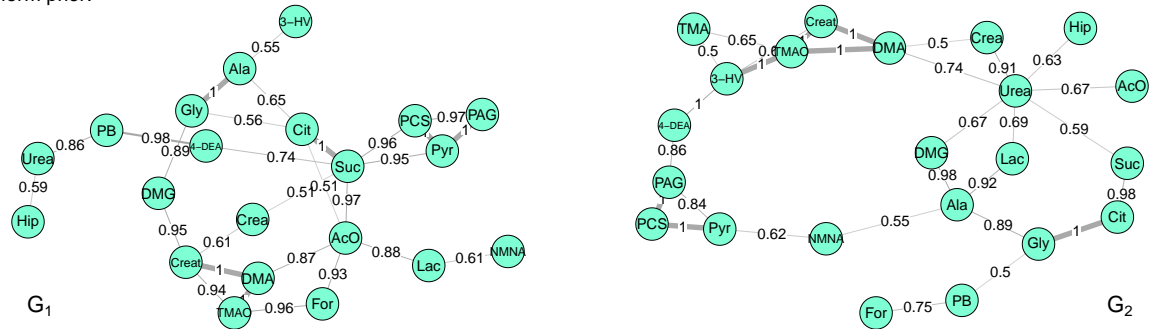


Figure S9: Plots of  $G_1$  and  $G_2$  showing edges with weights greater than 0.5 corresponding to different priors.

	In $G_1$ but not in $G_2$					In $G_2$ but not in $G_1$				
	Edge	00	11	10	01	Edge	00	11	10	01
$\sigma_1^2 = \sigma_2^2 = 1$ :	Pyr – Suc	0.01	0.01	0.98	0.00	3-HV – TMAO	0.00	0.01	0.00	0.99
	Suc – PCS	0.00	0.03	0.97	0.00	3-HV – 4-DEA	0.07	0.00	0.00	0.92
	AcO – PCS	0.40	0.00	0.59	0.01	Cit – Gly	0.00	0.15	0.00	0.85
	PCS – 4-DEA	0.38	0.03	0.59	0.01	TMA – TMAO	0.21	0.03	0.07	0.69
	DMG – Creat	0.43	0.01	0.56	0.00	Ala – DMG	0.41	0.01	0.00	0.59
	TMAO – For	0.43	0.01	0.56	0.01	DMA – Urea	0.45	0.01	0.05	0.50
	PB – 4-DEA	0.42	0.03	0.54	0.00	3-HV – Creat	0.51	0.01	0.00	0.48
	AcO – Pyr	0.45	0.01	0.53	0.01	Crea – Urea	0.47	0.00	0.05	0.47
	Pyr – 4-DEA	0.47	0.00	0.53	0.00	Ala – DMA	0.65	0.01	0.02	0.32
	AcO – Suc	0.49	0.00	0.51	0.00	AcO – DMA	0.32	0.01	0.37	0.31
$\sigma_1^2 = 1, \sigma_2^2 = 10$ :	Suc – PCS	0.00	0.01	0.98	0.00	3-HV – TMAO	0.00	0.00	0.00	1.00
	AcO – Suc	0.02	0.00	0.98	0.00	3-HV – 4-DEA	0.02	0.06	0.00	0.93
	Pyr – Suc	0.00	0.02	0.98	0.00	Cit – Gly	0.00	0.10	0.01	0.89
	PB – 4-DEA	0.27	0.00	0.73	0.00	3-HV – Creat	0.26	0.01	0.00	0.73
	Pyr – PAG	0.00	0.49	0.51	0.00	TMA – TMAO	0.46	0.03	0.01	0.49
	DMG – Creat	0.49	0.00	0.49	0.02	Ala – DMG	0.54	0.01	0.00	0.45
	Suc – 4-DEA	0.53	0.00	0.47	0.00	DMA – Urea	0.54	0.01	0.04	0.42
	Suc – Cit	0.00	0.54	0.43	0.03	PAG – 4-DEA	0.54	0.04	0.02	0.40
	Ala – Cit	0.57	0.04	0.38	0.01	Creat – TMAO	0.00	0.60	0.00	0.40
	Pyr – DMG	0.64	0.00	0.36	0.00	Lac – Ala	0.65	0.00	0.00	0.34
$\sigma_1^2 = \sigma_2^2 = 10$ :	AcO – Suc	0.07	0.00	0.93	0.00	3-HV – TMAO	0.00	0.02	0.00	0.98
	Suc – PCS	0.07	0.08	0.84	0.01	3-HV – 4-DEA	0.01	0.03	0.00	0.96
	Pyr – Suc	0.12	0.05	0.83	0.00	TMA – TMAO	0.23	0.05	0.00	0.71
	AcO – PCS	0.44	0.01	0.55	0.00	Cit – Gly	0.00	0.32	0.00	0.68
	AcO – Cit	0.45	0.01	0.54	0.00	3-HV – Creat	0.40	0.01	0.00	0.58
	Lac – AcO	0.53	0.01	0.47	0.00	Creat – TMAO	0.02	0.52	0.08	0.39
	Pyr – 4-DEA	0.51	0.01	0.45	0.03	PB – TMAO	0.64	0.00	0.00	0.36
	DMG – Creat	0.54	0.00	0.44	0.02	Ala – Pyr	0.63	0.00	0.01	0.35
	Suc – Cit	0.00	0.57	0.43	0.00	DMA – Urea	0.64	0.03	0.01	0.32
	Suc – 4-DEA	0.58	0.00	0.42	0.00	Creat – Gly	0.65	0.00	0.04	0.31
Uniform prior:	Suc – PCS	0.03	0.05	0.92	0.01	Ala – DMG	0.02	0.08	0.00	0.90
	AcO – For	0.07	0.02	0.91	0.00	3-HV – TMAO	0.00	0.10	0.00	0.90
	DMG – Creat	0.05	0.04	0.91	0.00	Lac – Ala	0.07	0.12	0.01	0.81
	AcO – Suc	0.03	0.10	0.87	0.00	PB – For	0.23	0.02	0.02	0.73
	Pyr – Suc	0.04	0.09	0.86	0.01	3-HV – 4-DEA	0.00	0.28	0.00	0.72
	DMG – Gly	0.10	0.04	0.85	0.01	Crea – Urea	0.07	0.26	0.01	0.66
	PB – Urea	0.11	0.05	0.81	0.03	AcO – Urea	0.27	0.04	0.06	0.63
	PB – 4-DEA	0.02	0.20	0.78	0.00	DMG – Urea	0.24	0.05	0.09	0.62
	AcO – DMA	0.11	0.12	0.75	0.02	3-HV – Creat	0.19	0.06	0.16	0.59
	TMAO – For	0.04	0.22	0.74	0.00	Ala – NMNA	0.42	0.01	0.03	0.54

Table S3: List of 10 edges which are most likely to appear in  $G_1$  but not in  $G_2$  (left) and in  $G_2$  but not in  $G_1$  (right) corresponding to the different priors. The posterior probability of each edge to appear in none of the graphs ('00'), both graphs ('11'), in  $G_1$  but not in  $G_2$  ('10') and in  $G_2$  but not in  $G_1$  ('01') are shown.



Node $i$	$\bar{\beta}_{i1}$	CI ( $\beta_{i1}$ )	$\bar{\beta}_{i2}$	CI ( $\beta_{i2}$ )	$\bar{\pi}_{1,i}$	CI ( $\pi_{1,i}$ )	$\bar{\pi}_{2,i}$	CI( $\pi_{2,i}$ )	$\bar{B}_{1,i}$	$\bar{B}_{2,i}$
TMAO	-0.44	(-1.69, 0.81)	0.30	(-1.27, 1.87)	0.40	(0.14, 0.68)	0.47	(0.14, 0.81)	0.07	0.18
PCS	-0.28	(-1.48, 0.94)	-0.50	(-1.96, 0.99)	0.44	(0.18, 0.71)	0.33	(0.07, 0.63)	0.10	0.07
Suc	-0.43	(-1.61, 0.75)	-0.91	(-2.41, 0.60)	0.40	(0.15, 0.67)	0.23	(0.03, 0.49)	0.18	0.02
DMA	-0.43	(-1.69, 0.84)	0.23	(-1.33, 1.81)	0.40	(0.14, 0.68)	0.46	(0.12, 0.80)	0.08	0.19
Creat	-0.63	(-1.88, 0.61)	0.08	(-1.50, 1.66)	0.36	(0.12, 0.62)	0.38	(0.08, 0.71)	0.06	0.12
4-DEA	-0.83	(-2.23, 0.55)	-0.43	(-2.02, 1.16)	0.32	(0.07, 0.60)	0.25	(0.02, 0.53)	0.08	0.04
Pyr	-0.23	(-1.44, 1.00)	-0.67	(-2.24, 0.90)	0.45	(0.18, 0.72)	0.31	(0.05, 0.62)	0.13	0.05
Cit	-0.76	(-2.04, 0.52)	-0.21	(-1.72, 1.31)	0.33	(0.09, 0.60)	0.30	(0.04, 0.61)	0.09	0.11
3-HV	-1.39	(-2.65, -0.14)	0.28	(-1.22, 1.79)	0.22	(0.05, 0.42)	0.27	(0.04, 0.55)	0.00	0.08
Gly	-0.95	(-2.23, 0.32)	-0.12	(-1.70, 1.46)	0.30	(0.07, 0.55)	0.28	(0.03, 0.57)	0.05	0.11
Urea	-1.27	(-2.67, 0.11)	-0.07	(-1.76, 1.59)	0.24	(0.04, 0.48)	0.24	(0.01, 0.55)	0.02	0.07
Ala	-0.98	(-2.29, 0.32)	-0.19	(-1.85, 1.46)	0.29	(0.07, 0.54)	0.27	(0.02, 0.59)	0.05	0.10
PAG	-0.85	(-2.09, 0.38)	-0.24	(-1.83, 1.33)	0.31	(0.09, 0.56)	0.28	(0.03, 0.58)	0.02	0.06
AcO	-0.74	(-2.13, 0.65)	-0.89	(-2.52, 0.75)	0.34	(0.09, 0.62)	0.20	(0.01, 0.46)	0.12	0.02
Hip	-1.65	(-3.08, -0.24)	-0.62	(-2.28, 1.05)	0.18	(0.02, 0.39)	0.12	(0.00, 0.32)	0.01	0.01
DMG	-1.19	(-2.62, 0.21)	-0.62	(-2.26, 0.99)	0.25	(0.04, 0.50)	0.17	(0.00, 0.41)	0.06	0.01
TMA	-1.42	(-2.79, -0.06)	-0.37	(-1.96, 1.20)	0.22	(0.04, 0.43)	0.17	(0.01, 0.39)	0.02	0.02
Lac	-1.66	(-3.08, -0.24)	-0.61	(-2.28, 1.05)	0.18	(0.02, 0.39)	0.13	(0.00, 0.35)	0.01	0.01
PB	-1.39	(-2.77, -0.03)	-0.47	(-2.18, 1.18)	0.22	(0.04, 0.44)	0.17	(0.00, 0.41)	0.02	0.01
NMNA	-1.52	(-2.91, -0.14)	-0.58	(-2.28, 1.09)	0.20	(0.03, 0.41)	0.14	(0.00, 0.37)	0.01	0.01
For	-1.40	(-2.85, -0.01)	-0.70	(-2.33, 0.92)	0.22	(0.03, 0.44)	0.14	(0.00, 0.34)	0.03	0.01
Crea	-1.42	(-2.87, 0.03)	-0.48	(-2.18, 1.16)	0.22	(0.03, 0.46)	0.17	(0.00, 0.42)	0.02	0.02

Table S4: Weighted mean ( $\bar{\pi}_{k,i}$ ) and 95% credible interval (CI) of  $\pi_{k,i}$ , weighted mean ( $\bar{\beta}_{iq}$ ) and 95% credible interval (CI) of  $\beta_{iq}$ , and weighted mean betweenness centrality ( $\bar{B}_{k,i}$ ) for each node  $i$  for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .

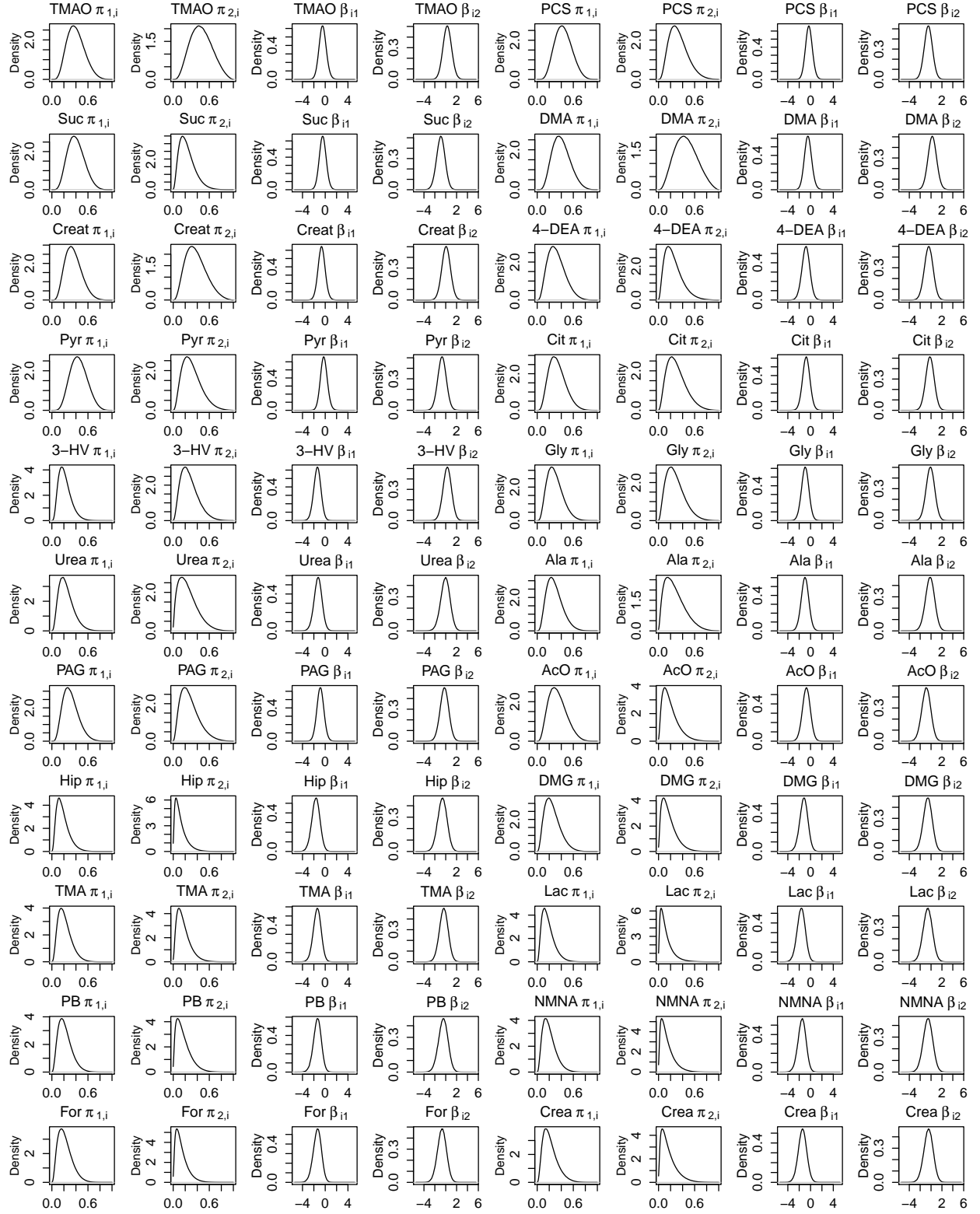


Figure S10: Posterior distributions of the connectivity  $\pi_{k,i}$  and the regression coefficients  $\beta_{iq}$  for each metabolite for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .

Node $i$	$\bar{\beta}_{i1}$	CI ( $\beta_{i1}$ )	$\bar{\beta}_{i2}$	CI ( $\beta_{i2}$ )	$\bar{\pi}_{1,i}$	CI ( $\pi_{1,i}$ )	$\bar{\pi}_{2,i}$	CI( $\pi_{2,i}$ )	$\bar{B}_{1,i}$	$\bar{B}_{2,i}$
TMAO	-0.76	(-2.16, 0.64)	0.71	(-1.96, 3.51)	0.34	(0.08, 0.62)	0.47	(0.07, 0.92)	0.04	0.10
PCS	-0.38	(-1.73, 1.00)	-0.55	(-3.01, 1.86)	0.41	(0.14, 0.71)	0.31	(0.02, 0.69)	0.08	0.07
Suc	-0.04	(-1.37, 1.34)	-2.46	(-6.03, 0.61)	0.49	(0.20, 0.79)	0.14	(0.00, 0.11)	0.24	0.02
DMA	-0.64	(-1.97, 0.70)	0.48	(-2.35, 3.31)	0.36	(0.11, 0.64)	0.45	(0.04, 0.92)	0.06	0.12
Creat	-0.61	(-2.03, 0.81)	0.48	(-2.28, 3.37)	0.37	(0.09, 0.67)	0.46	(0.06, 0.94)	0.08	0.09
4-DEA	-0.80	(-2.24, 0.64)	-0.50	(-3.14, 1.99)	0.33	(0.07, 0.62)	0.26	(0.00, 0.61)	0.06	0.05
Pyr	-0.17	(-1.56, 1.28)	-1.20	(-3.80, 1.33)	0.46	(0.16, 0.77)	0.25	(0.00, 0.61)	0.13	0.03
Cit	-0.83	(-2.24, 0.59)	-0.78	(-3.53, 1.86)	0.32	(0.07, 0.61)	0.22	(0.00, 0.56)	0.07	0.05
3-HV	-1.39	(-2.82, 0.02)	0.98	(-1.69, 3.70)	0.22	(0.03, 0.45)	0.41	(0.04, 0.84)	0.00	0.09
Gly	-0.99	(-2.37, 0.38)	-0.27	(-3.07, 2.39)	0.29	(0.06, 0.56)	0.27	(0.00, 0.64)	0.04	0.08
Urea	-1.35	(-2.82, 0.08)	-0.98	(-5.71, 2.78)	0.23	(0.03, 0.47)	0.18	(0.00, 0.15)	0.02	0.04
Ala	-0.93	(-2.31, 0.45)	-0.27	(-3.73, 2.98)	0.30	(0.07, 0.57)	0.29	(0.00, 0.76)	0.05	0.09
PAG	-0.84	(-2.20, 0.53)	-0.26	(-3.21, 2.52)	0.32	(0.08, 0.59)	0.30	(0.00, 0.74)	0.03	0.06
AcO	-0.79	(-2.24, 0.66)	-3.25	(-7.22, 0.28)	0.33	(0.07, 0.63)	0.05	(0.00, 0.05)	0.09	0.00
Hip	-1.47	(-3.01, 0.08)	-2.89	(-6.96, 0.81)	0.21	(0.03, 0.46)	0.04	(0.00, 0.04)	0.02	0.00
DMG	-1.12	(-2.63, 0.36)	-1.84	(-5.84, 1.60)	0.27	(0.04, 0.54)	0.12	(0.00, 0.10)	0.05	0.01
TMA	-1.55	(-2.97, -0.17)	-0.74	(-3.86, 2.19)	0.20	(0.03, 0.41)	0.15	(0.00, 0.52)	0.00	0.02
Lac	-1.35	(-2.80, 0.06)	-1.01	(-4.92, 2.26)	0.23	(0.03, 0.47)	0.16	(0.00, 0.14)	0.01	0.03
PB	-1.23	(-2.65, 0.17)	-2.82	(-6.89, 0.87)	0.25	(0.04, 0.49)	0.06	(0.00, 0.05)	0.02	0.00
NMNA	-1.33	(-2.86, 0.17)	-1.96	(-6.03, 1.53)	0.23	(0.03, 0.48)	0.10	(0.00, 0.08)	0.01	0.01
For	-1.41	(-2.84, 0.02)	-1.74	(-5.84, 1.73)	0.22	(0.03, 0.45)	0.11	(0.00, 0.09)	0.01	0.01
Crea	-1.40	(-2.86, 0.04)	-2.68	(-6.82, 1.07)	0.22	(0.03, 0.46)	0.06	(0.00, 0.05)	0.02	0.00

Table S5: Weighted mean ( $\bar{\pi}_{k,i}$ ) and 95% credible interval (CI) of  $\pi_{k,i}$ , weighted mean ( $\bar{\beta}_{iq}$ ) and 95% credible interval (CI) of  $\beta_{iq}$ , and weighted mean betweenness centrality ( $\bar{B}_{k,i}$ ) for each node  $i$  for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = 1$ ,  $\sigma_2^2 = 10$ .

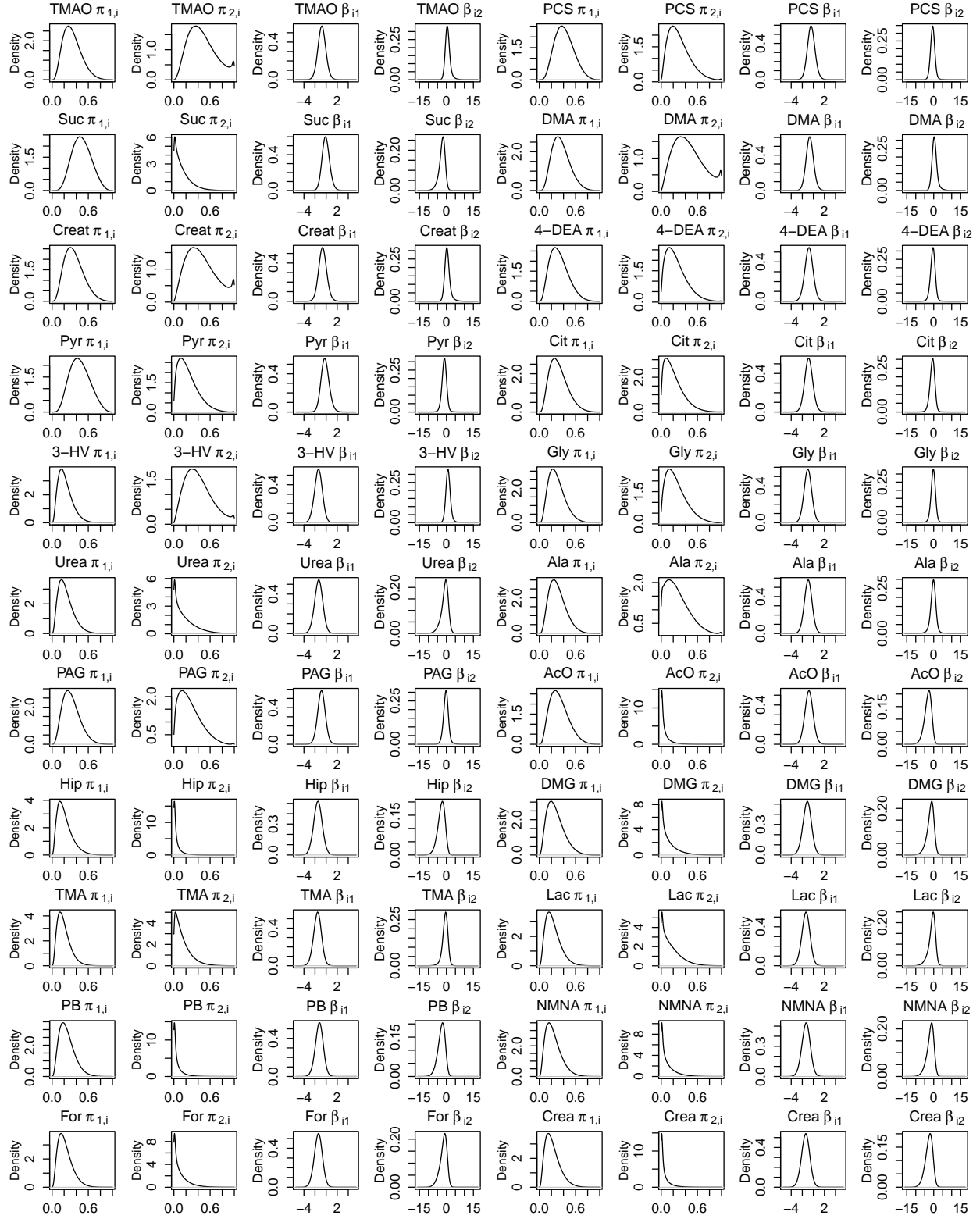


Figure S11: Posterior distributions of the connectivity  $\pi_{k,i}$  and the regression coefficients  $\beta_{iq}$  for each metabolite for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = 1$  and  $\sigma_2^2 = 10$ .

Node $i$	$\bar{\beta}_{i1}$	CI ( $\beta_{i1}$ )	$\bar{\beta}_{i2}$	CI ( $\beta_{i2}$ )	$\bar{\pi}_{1,i}$	CI ( $\pi_{1,i}$ )	$\bar{\pi}_{2,i}$	CI( $\pi_{2,i}$ )	$\bar{B}_{1,i}$	$\bar{B}_{2,i}$
TMAO	-1.09	(-3.33, 1.02)	1.63	(-1.59, 5.13)	0.29	(0.01, 0.66)	0.58	(0.00, 0.90)	0.03	0.14
PCS	0.16	(-2.26, 3.02)	-0.69	(-4.25, 2.86)	0.52	(0.15, 0.99)	0.38	(0.00, 0.90)	0.10	0.05
Suc	-0.07	(-2.12, 2.09)	-2.48	(-6.31, 0.87)	0.48	(0.11, 0.89)	0.14	(0.00, 0.12)	0.12	0.01
DMA	-0.97	(-2.90, 0.88)	0.82	(-2.19, 3.86)	0.30	(0.02, 0.66)	0.45	(0.03, 0.92)	0.05	0.09
Creat	-0.91	(-3.55, 1.74)	0.43	(-3.52, 4.39)	0.33	(0.00, 0.78)	0.42	(0.05, 0.99)	0.07	0.06
4-DEA	-1.59	(-4.62, 0.95)	-0.38	(-3.72, 3.00)	0.23	(0.00, 0.63)	0.18	(0.00, 0.53)	0.03	0.01
Pyr	0.12	(-2.19, 2.74)	-0.45	(-4.18, 3.20)	0.51	(0.14, 0.96)	0.42	(0.02, 0.94)	0.10	0.07
Cit	-0.88	(-3.55, 1.66)	-0.01	(-3.32, 3.26)	0.33	(0.00, 0.76)	0.33	(0.00, 0.85)	0.08	0.07
3-HV	-2.95	(-5.83, -0.33)	2.14	(-1.12, 5.59)	0.09	(0.00, 0.07)	0.34	(0.01, 0.75)	0.00	0.08
Gly	-1.30	(-3.55, 0.81)	0.30	(-2.92, 3.40)	0.26	(0.00, 0.60)	0.31	(0.00, 0.72)	0.06	0.07
Urea	-3.41	(-7.04, -0.33)	0.37	(-5.25, 5.06)	0.08	(0.00, 0.06)	0.17	(0.00, 0.17)	0.00	0.03
Ala	-1.75	(-3.90, 0.24)	0.69	(-3.78, 4.73)	0.19	(0.00, 0.48)	0.35	(0.00, 0.32)	0.02	0.09
PAG	-1.12	(-3.19, 0.88)	0.05	(-3.05, 3.13)	0.28	(0.01, 0.64)	0.30	(0.00, 0.76)	0.02	0.04
AcO	-0.55	(-3.26, 2.09)	-2.70	(-6.71, 0.94)	0.40	(0.00, 0.83)	0.10	(0.00, 0.08)	0.10	0.00
Hip	-3.75	(-7.40, -0.62)	-0.75	(-6.31, 4.33)	0.06	(0.00, 0.05)	0.07	(0.00, 0.06)	0.00	0.00
DMG	-2.04	(-5.47, 0.67)	-1.55	(-6.78, 3.40)	0.19	(0.00, 0.16)	0.09	(0.00, 0.08)	0.05	0.01
TMA	-3.19	(-6.19, -0.55)	0.51	(-3.85, 4.73)	0.08	(0.00, 0.06)	0.13	(0.00, 0.11)	0.00	0.01
Lac	-2.51	(-6.33, 0.52)	-1.19	(-6.38, 3.46)	0.15	(0.00, 0.12)	0.10	(0.00, 0.08)	0.01	0.01
PB	-3.14	(-6.76, -0.12)	0.26	(-5.11, 5.06)	0.09	(0.00, 0.07)	0.17	(0.00, 0.15)	0.00	0.02
NMNA	-3.30	(-7.19, -0.12)	-1.31	(-6.58, 3.46)	0.09	(0.00, 0.07)	0.06	(0.00, 0.06)	0.00	0.00
For	-2.87	(-6.69, 0.17)	-0.74	(-6.05, 3.99)	0.12	(0.00, 0.09)	0.12	(0.00, 0.10)	0.01	0.01
Crea	-3.18	(-7.04, 0.31)	-0.89	(-6.58, 4.46)	0.11	(0.00, 0.07)	0.09	(0.00, 0.07)	0.01	0.01

Table S6: Weighted mean ( $\bar{\pi}_{k,i}$ ) and 95% credible interval (CI) of  $\pi_{k,i}$ , weighted mean ( $\bar{\beta}_{iq}$ ) and 95% credible interval (CI) of  $\beta_{iq}$ , and weighted mean betweenness centrality ( $\bar{B}_{k,i}$ ) for each node  $i$  for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 10$ .

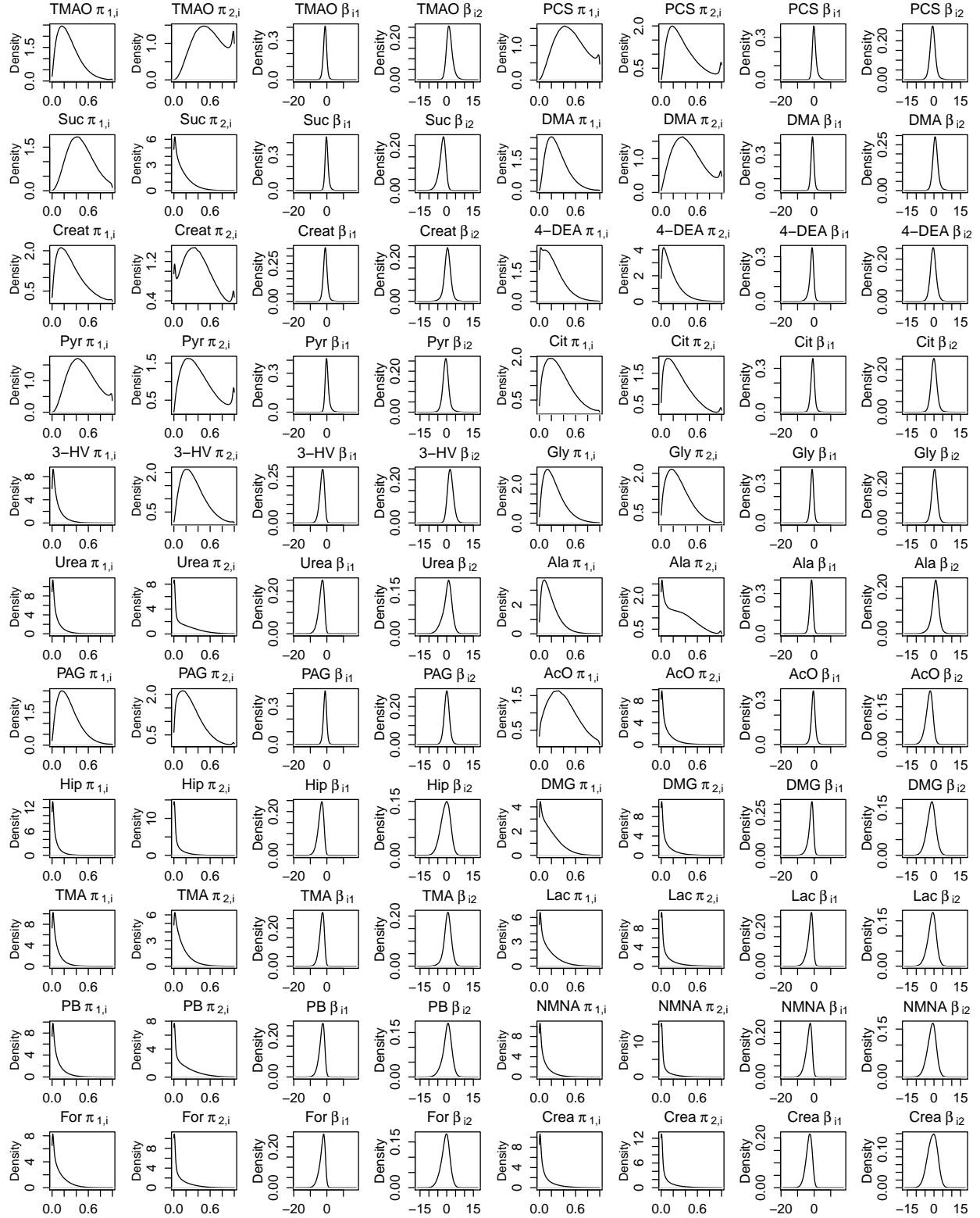


Figure S12: Posterior distributions of the connectivity  $\pi_{k,i}$  and the regression coefficients  $\beta_{iq}$  for each metabolite for  $k = 1, 2$  and  $q = 1, 2$  corresponding to the multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 10$ .

	TMAO	PCS	Suc	DMA	Creat	4-DEA	Pyr	Cit	3-HV	Gly	Urea	Ala	PAG	AcO	Hip	DMG	TMA	Lac	PB	NMNA	For	Crea
TMAO	1.74	0.00	0.01	-1.09	0.44	0.01	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.01	0.06	0.00	-0.00	-0.03	0.00	-0.00	-0.00	-0.24	0.00
PCS	.	4.14	-1.22	0.00	0.01	-0.40	-2.54	0.00	0.00	0.00	-0.00	0.00	-0.87	-0.49	0.00	0.04	0.00	0.03	0.00	0.04	-0.00	0.00
Suc	.	.	2.19	0.01	0.00	-0.13	1.06	-0.72	-0.00	-0.00	0.00	0.00	0.02	-0.35	-0.00	0.00	0.01	0.00	-0.00	0.01	-0.05	-0.07
DMA	.	.	.	2.34	-1.14	-0.00	0.00	-0.00	0.00	0.00	0.01	0.01	0.00	0.15	0.00	-0.01	-0.02	0.00	-0.00	0.01	-0.00	0.00
Creat	.	.	.	.	1.80	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	-0.25	-0.00	0.00	-0.00	0.01	0.00	-0.02
4-DEA	.	.	.	.	.	1.39	0.36	-0.00	-0.00	-0.00	0.00	0.00	0.01	-0.00	-0.00	0.00	-0.00	0.00	-0.28	-0.00	-0.00	-0.08
Pyr	.	.	.	.	.	.	4.29	-0.01	0.00	-0.00	-0.01	-0.00	-1.40	0.39	0.00	0.11	0.01	0.01	-0.00	0.09	0.00	0.05
Cit	.	.	.	.	.	.	.	1.56	-0.00	-0.05	0.04	-0.27	-0.00	0.12	-0.00	0.02	0.01	0.00	-0.00	0.00	-0.00	0.03
3-HV	.	.	.	.	.	.	.	.	1.06	-0.00	-0.00	-0.03	0.00	0.00	0.00	-0.01	-0.00	0.00	0.00	-0.00	0.00	0.00
Gly	.	.	.	.	.	.	.	.	.	1.53	0.00	-0.73	-0.00	0.00	0.01	-0.16	-0.00	-0.00	0.00	0.00	-0.01	0.00
Urea	.	.	.	.	.	.	.	.	.	.	1.13	0.00	-0.00	-0.00	0.02	-0.00	-0.02	0.01	0.11	0.00	0.00	0.01
Ala	.	.	.	.	.	.	.	.	.	.	.	1.57	0.00	0.00	0.02	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
PAG	.	.	.	.	.	.	.	.	.	.	.	.	2.74	0.00	0.00	0.00	0.01	0.01	-0.00	0.01	0.00	0.00
AcO	.	.	.	.	.	.	.	.	.	.	.	.	.	1.59	-0.00	0.00	-0.00	-0.11	0.01	0.01	-0.20	-0.01
Hip	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.08	0.00	0.00	0.00	0.00	-0.02	-0.02	-0.00
DMG	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.27	-0.02	0.00	0.00	-0.00	0.00	-0.01
TMA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.09	-0.00	0.01	-0.00	-0.00	0.01
Lac	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.11	0.00	0.03	0.00	0.00
PB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.22	0.00	0.00	0.00
NMNA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.12	-0.00	-0.00
For	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.23	-0.00
Crea	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.13

Table S7: Mean precision matrix  $\Omega_1$  corresponding to  $K = 2$  and multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .

	TMAO	PCS	Suc	DMA	Creat	4-DEA	Pyr	Cit	3-HV	Gly	Urea	Ala	PAG	AcO	Hip	DMG	TMA	Lac	PB	NMNA	For	Crea
TMAO	5.23	-0.01	0.00	-3.27	1.38	-0.00	0.02	0.00	-1.85	-0.01	0.01	-0.02	-0.01	0.01	0.01	-0.00	-0.45	-0.00	-0.18	0.01	0.00	-0.03
PCS	.	11.46	-0.00	-0.02	0.00	0.00	-8.09	-0.01	0.00	0.00	-0.00	-0.00	-3.53	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.01	-0.00	0.00
Suc	.	.	1.30	-0.00	-0.03	0.00	-0.00	-0.48	0.00	-0.01	0.04	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00
DMA	.	.	.	4.58	-1.56	-0.06	-0.00	-0.00	-0.03	-0.04	0.22	-0.11	-0.01	0.08	0.06	-0.00	-0.08	-0.00	0.12	-0.00	0.04	-0.04
Creat	.	.	.	.	1.86	-0.05	0.00	-0.06	-0.33	-0.03	0.00	-0.00	0.01	0.01	0.04	0.00	-0.00	0.00	0.03	0.00	0.06	0.00
4-DEA	.	.	.	.	.	1.57	0.00	-0.00	-0.75	-0.00	0.00	-0.01	0.13	0.00	0.00	-0.00	0.00	-0.01	-0.01	0.00	0.00	0.00
Pyr	.	.	.	.	.	.	8.02	-0.01	0.00	-0.00	-0.01	-0.02	0.84	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.06	0.00	0.00
Cit	.	.	.	.	.	.	.	1.83	-0.00	-0.87	0.00	-0.04	-0.01	0.01	0.00	-0.02	0.00	0.00	0.04	0.00	0.00	-0.00
3-HV	.	.	.	.	.	.	.	.	3.03	-0.01	0.02	-0.01	0.00	0.01	0.00	-0.00	-0.05	-0.00	-0.00	-0.03	0.00	-0.00
Gly	.	.	.	.	.	.	.	.	.	1.74	0.07	-0.27	0.01	0.00	0.00	-0.00	0.01	-0.01	-0.04	0.00	-0.00	-0.00
Urea	.	.	.	.	.	.	.	.	.	.	1.33	0.03	-0.00	-0.04	0.02	-0.04	0.00	0.05	-0.00	0.03	0.00	-0.21
Ala	.	.	.	.	.	.	.	.	.	.	.	1.41	0.01	0.01	0.02	-0.30	-0.00	-0.07	0.00	0.02	-0.00	-0.00
PAG	.	.	.	.	.	.	.	.	.	.	.	.	3.43	0.00	-0.00	0.00	-0.00	0.00	0.01	0.01	-0.00	-0.00
AcO	.	.	.	.	.	.	.	.	.	.	.	.	.	1.10	0.00	0.00	0.00	-0.00	-0.01	0.00	0.00	-0.00
Hip	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.10	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00
DMG	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.20	-0.00	-0.00	-0.00	0.00	-0.00	-0.03
TMA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.34	0.00	-0.00	-0.00	0.00	0.00
Lac	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.10	-0.00	0.00	-0.00	-0.00
PB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.15	0.00	-0.04	-0.00
NMNA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.10	-0.00	0.02
For	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.10	0.01
Crea	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.16

Table S8: Mean precision matrix  $\Omega_2$  corresponding to  $K = 2$  and multiplicative prior with  $\sigma_1^2 = \sigma_2^2 = 1$ .